1941

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 42

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

Annals of Mathematics
Vol. 42, No. 1, January, 1941

ON FROBENIUSEAN ALGEBRAS. II

By Tadasi Nakayama

(Received July 13, 1939)

Introduction

In the present paper we continue our study of a class of (associative) algebras called Frobeniusean.[1] Chapter I, which is short, is a direct sequel of Part I and deals chiefly with a certain type of automorphisms in a Frobeniusean algebra. The automorphisms enable us to generalize and refine some of the theorems in Part I, and also clarify the significance of symmetric algebras. Further, an application of our results to the theory of Galois moduli and normal bases over modular fields is discussed. In Chapter II we make an attempt to extend our theory from algebras to general rings satisfying chain conditions. To do so we adopt the properties of I, 2, Lemma 2 as a definition of Frobeniusean and quasi-Frobeniusean rings. Then many of the theorems in I can be extended to this general case. But since this definition does not have much significance for itself, contrary to the case of algebras, Chapter II may be considered as a study of the structure of those rings in which the annihilation gives a 1-1 correspondence between left and right ideals.[2] Sections 6, 7 aim, however, not only at Frobeniusean rings but at Frobeniusean algebras.

Chapter I: A Further Study of Frobeniusean Algebras

1. A class of automorphisms in a Frobeniusean algebra

Let A be a Frobeniusean algebra over a field F. Let $a_1, a_2, \cdots, a_n$ be a basis of A with a multiplication table

$$a_\rho a_\sigma = \sum_\tau \alpha_{\rho\sigma\tau} a_\tau . \tag{1}$$

There exists a vector (λ_τ) such that the corresponding parastrophic matrix $P = (\sum_\tau \alpha_{\rho\sigma\tau}\lambda_\tau)$ is non-singular. We consider again the linear form $\lambda(x) = \sum \lambda_\rho \xi_\rho$ $(x = \sum \xi_\rho a_\rho \in A)$ and the *non-singular* hyperplane H defined by $\lambda(x) = 0$. For a set S of elements in A we denote by $p_H(S)[q_H(S)]$, as before, the set of all x satisfying $xS \subseteq H[Sx \subseteq H]$. Any other non-singular hyperplane H' in A can be given as $H' = p_H(c^{-1}) = Hc$ with a regular element c, and conversely with any regular element c Hc is a non-singular hyperplane in A (Cf. I, 6).

[1] *On Frobeniusean algebras. I*, Ann. Math. 40 (1939) referred to as Part I, or simply as I.
[2] M. Hall (7) calls them (weakly) closed rings.

THEOREM 1. *There exists a uniquely determined automorphism* $\varphi: x \to x^*$ *of* A *such that*

$$(2) \qquad \lambda(x^*y) = \lambda(yx) \qquad (\textit{that is, } x^*y - yx \in H)$$

for all $y \in A$. H *remains invariant under* φ; $H^* = H$.

φ*'s corresponding to different choices of non-singular hyperplanes in* A *form a co-set with respect to the invariant subgroup consisting of all inner automorphisms in the automorphism group of* A.

PROOF. If $x = \sum \xi_\rho a_\rho$ we put $(\xi_\rho^*) = P(P')^{-1}(\xi_\rho)$ and $x^* = \sum \xi_\rho^* a_\rho$. Then[3] $\lambda(x^*y) = (\xi^*)P(\eta) = (\xi)P'(\eta) = (\eta)P(\xi) = \lambda(yx)$ where $y = \sum \eta_\rho a_\rho$. This shows the existence of a mapping $\varphi: x \to x^*$ such that (2) is true for all $x, y \in A$. φ is uniquely determined by this property as we can see easily by reversing the above computation and observing particularly that P is non-singular. The mapping is evidently 1-1, and that $(\alpha x + \beta z)^* = \alpha x^* + \beta z^*$ is obvious. Moreover, the relation $\lambda(x^*z^*y) = \lambda(z^*yx) = \lambda(yxz)$ shows that $(xz)^* = x^*z^*$. Hence φ is an automorphism. Now, $x \in H$ implies $x^* \in H$, since $x^* - x = x^*1 - 1x \in H$. Thus $H^* = H$.

To prove the second part of the theorem, consider a second non-singular hyperplane $H' = Hc$ (c regular). Since $(cxc^{-1})^*y - yx = ((cxc^{-1})^*yc^{-1} - yc^{-1}(cxc^{-1}))c \in Hc = H'$ for all $x, y \in A$, the automorphism φ' corresponding to H' is given by $x \to (cxc^{-1})^*$.

LEMMA 1. *The automorphism* φ *maps* $q_H(S)$, *with any set* S *in* A, *onto* $p_H(S)$; $q_H(S)^* = p_H(S)$. *In particular* $r(\mathfrak{l})^*(= r(\mathfrak{l}^*)) = p_H(\mathfrak{l})$ *for a left ideal* $\mathfrak{l}$ *in* A.[4]

PROOF. Immediate from the definition and I, Lemma 4.

Thus the automorphism φ *correlates the two kinds of dual correspondences between left and right ideals in* A; one is given by annihilation and the other is representation-theoretical. Namely

THEOREM 2.[5] *If* $\mathfrak{l}, \mathfrak{l}_0$ *are left ideals in* A *and* $\mathfrak{l} \supseteqq \mathfrak{l}_0$, *the representation of* A *defined by the left module* $\mathfrak{l}/\mathfrak{l}_0$ *is equivalent to the one defined by the right module* $r(\mathfrak{l}_0)^*/r(\mathfrak{l})^*$. *Furthermore, the representation defined by a left principal ideal* Ac *is equivalent to the one defined by the right ideal* c^*A.

PROOF. The theorem follows from the above lemma and I, 8, Lemma 6. For the second half cf. also the proof of I, 5, Th. 4.

THEOREM 3. *The automorphism algebra of the left module* $\mathfrak{l}/\mathfrak{l}_0[Ac]$ *is inversely isomorphic to that of the right module* $r(\mathfrak{l}_0)/r(\mathfrak{l})[cA]$.[6]

PROOF. The automorphism algebra of a left [right] module is inversely [directly] isomorphic to the algebra of matrices which commute with all the matrices in a representation belonging to the module. It follows thus from Theorem 2 that the automorphism algebra of $\mathfrak{l}/\mathfrak{l}_0$ is inversely isomorphic to

[3] Consult the form of the parastrophic matrix P. Cf. also I, 3.

[4] $r(S)$ denotes, as before, the set of right annihilators of S in A.

[5] Cf. Th. 13 in the section 6 below.

[6] The last statement of I, Th. 5 is a very special case of the present theorem.

that of $r(\mathfrak{l}_0)^*/r(\mathfrak{l})^*$. But this latter is evidently isomorphic to the automorphism algebra of $r(\mathfrak{l}_0)/r(\mathfrak{l})$.

Now, let E_κ, $e_{\kappa,i}$ ($\kappa = 1, 2, \cdots, k$; $i = 1, 2, \cdots, f(\kappa)$) and $\pi(\kappa)$ ($\kappa = 1, 2, \cdots, k$) have the same significance as in I. In particular $U_\kappa \cong V_{\pi(\kappa)}$. Then

THEOREM 4. $E^*_{\pi(\kappa)} = E_\kappa$. *With a suitable choice of H we have furthermore $e^*_{\pi(\kappa),i} = e_{\kappa,i}$ (for all κ, i).*[7]

PROOF. Obviously $E^*_{\pi(\kappa)} = E_\lambda$ for a certain λ. But the representation defined by $E_\lambda A$ must be equivalent to the one defined by $AE_{\pi(\kappa)}$. Hence $\lambda = \kappa$. To prove the second half we observe that we can choose H so that $e_{\kappa,i}Ae_{\lambda,j} \subseteq H$ if $(\pi(\kappa), i) \neq (\lambda, j)$; this can be seen in exactly the same way as in Nakayama-Nesbitt (11). With such a choice of H we have indeed $e^*_{\pi(\kappa),i} = e_{\kappa,i}$. For, $\lambda(e_{\kappa,i}x) = \lambda(\sum_{\lambda,j} e_{\kappa,i}xe_{\lambda,j}) = \lambda(e_{\kappa,i}xe_{\pi(\kappa),i}) = \lambda(\sum_{\lambda,j} e_{\lambda,j}xe_{\pi(\kappa),i}) = \lambda(xe_{\pi(\kappa),i})$.

Another immediate consequence of Lemma 1 is that if $\mathfrak{z}$ is a two-sided ideal in A then

$$l(\mathfrak{z}) = r(\mathfrak{z})^* = r(\mathfrak{z}^*).$$

In particular if $\mathfrak{z} = \mathfrak{z}^*$, or if more particularly $\mathfrak{z}$ is invariant under all automorphisms of A, then $l(\mathfrak{z}) = r(\mathfrak{z})$. (This proves again I, Th. 6 in case where the algebra is not only quasi-Frobeniusean but Frobeniusean.) We find moreover that we can put $d = c^*$ in I, Th. 9, Corollary.

We note further that if $S(x)$ and $R(x)$ denote, as before, the left and the right regular representations of A with respect to the basis $a_1, a_2, \cdots, a_n$ then we have the relation

$$R(x)P' = P'S(x^*)$$

as a counterpart of the fundamental relation $R(x)P = PS(x)$. For, $(a_1^*, \cdots, a_n^*) = (a_1, \cdots, a_n)(P')^{-1}P$, as can be seen from the proof of Theorem 1, and if we put $\tilde{R}(x) = R(\varphi^{-1}(x))$ then $x \rightarrow \tilde{R}(x)$ is the right regular representation with respect to the basis (a_i^*), whence $\tilde{R}(x) = P'P^{-1}R(x)(P'P^{-1})^{-1} = P'S(x)(P')^{-1}$. But this is nothing but the above relation.

For an interesting application of the automorphism φ to a generalization of the orthogonality relation among representation coefficients, see a forthcoming paper by C. Nesbitt.

REMARK. Our algebra A is symmetric if and only if the automorphism φ is inner. The above treatment of Frobeniusean algebras can be considered as an extension of the section 9 in I.

EXAMPLE. *Cartan's algebra of outer multiplication*[8] is Frobeniusean. Namely, let m be any natural number and consider an algebra A over a field F which is generated by F and m elements $c_1, c_2, \cdots, c_m$ and in which the law of compo-

[7] The automorphism φ^{-1} effects the permutation π, so to speak.

[8] See for instance Cartan (19). This example was suggested to the writer by C. Chevalley.

sition is given by $c_i^2 = 0$, $c_ic_j = -c_jc_i$ for $i \neq j$. This algebra A, of Cartan's outer multiplication, has a basis 1, $c_{i_1}c_{i_2} \cdots c_{i_s}$ ($i_1 < i_2 < \cdots < i_s$), and is Frobeniusean. For, if H is the hyperplane in A consisting of those elements which have a vanishing coefficient for the term $c_1c_2 \cdots c_m$ when expressed by the above basis, then evidently H does not contain any ideal except the zero ideal. The automorphism φ belonging to the same H is defined simply by $c_i \to c_i^* = (-1)^{m-1}c_i$. If the characteristic of F is different from 2, then A is symmetric or not according as m is odd or even.[9]

2. Galois moduli over modular fields

Let K be a (finite and separable) Galois extension over F, and let $\mathfrak{G}$ be the Galois group of K/F. Looking upon the elements of $\mathfrak{G}$ as the right operators on K, we can consider K/F as a *right* representation module of the group algebra $\mathfrak{G}(F)$ of $\mathfrak{G}$ over F. The well-known theorem of normal bases[10] states that this module is (operator-) isomorphic with $\mathfrak{G}(F)$ itself, that is, K/F defines a representation of $\mathfrak{G}$ equivalent to the regular representation. Furthermore, the image of a *left* ideal of $\mathfrak{G}(F)$ by such an (operator-) isomorphism between $\mathfrak{G}(F)$ and K/F is independent of the special choice of isomorphism, and such an image of a left ideal of $\mathfrak{G}(F)$ in K is called a *Galois module* of K/F.[11]

Let K_1 be a field between F and K; $K \supseteqq K_1 \supseteqq F$, and let $\mathfrak{H}$ be the subgroup of $\mathfrak{G}$ belonging to K_1. The group algebra $\mathfrak{H}(K_1)$ of $\mathfrak{H}$ over K_1 is then (operator-) isomorphic to K/K_1 with respect to the right operator algebra $\mathfrak{H}(K_1)$, and Galois moduli of K/K_1 are defined in the same way as above. Now, the following interesting theorem was proved in Deuring (21) under the assumption that the group algebra $\mathfrak{G}(F)$ is semisimple or, what is equivalent, that the degree $(K:F)$ is not divisible by the characteristic of F:

THEOREM 5. *Let* $\mathfrak{m}$ *be a Galois module of* K/F *such that it is a* K_1*-module. Then* $\mathfrak{m}$ *is also a Galois module of* K/K_1*, and moreover, the representation of* $\mathfrak{G}$ *obtained from the left ideal of* $\mathfrak{G}(F)$ *corresponding to the* K/F*-Galois-module* $\mathfrak{m}$ *is equivalent to the one induced from the representation of the subgroup* $\mathfrak{H}$ *obtained from the left ideal of* $\mathfrak{H}(K_1)$ *corresponding to the* K/K_1*-Galois-module* $\mathfrak{m}$.

The purpose of the present section is to note that this theorem holds *without assuming the semisimplicity of* $\mathfrak{G}(F)$.

PROOF. Let $\mathfrak{l}$ be the left ideal of $\mathfrak{G}(F)$ which corresponds to our K/F-Galois-

[9] A can be considered also as the algebra of chains contained in an (absolute) simplex spanned by m vertices $c_1, c_2, \cdots, c_m$.

[10] Noether (26), Deuring (20), (21), Brauer (18). Artin has given another simple and elegant proof.

[11] Deuring (21). It seems to the writer that there is a slight confusion in the usage of the term "Galois module." Noether defined first a Galois module as an image of a right ideal. Deuring used *essentially* the same definition in his paper (20), but switched in his second paper (21), as well as in his book (5), to the definition which we are adopting in the present paper.

module $\mathfrak{m}$. $\mathfrak{l}$ consists of all left annihilators of $\mathfrak{r} = r(\mathfrak{l})$ (I, Th. 1. Cf. also I, 9), and thus Deuring's method can be transferred term by term to our general case. Namely, $\mathfrak{m}$ consists of all elements a in K satisfying

$$\sum_S a^S \beta_S = 0 \quad \text{for all} \quad \sum_S S\beta_S \in \mathfrak{r}. \tag{3}$$

Due to our assumption that $\mathfrak{m}$ is a K_1-module, we have

$$\sum_S h^S a^S \beta_S = \sum_T h^T \sum_{P \in \mathfrak{H}} a^{PT} \beta_{PT} = 0$$

for any h in K_1, where $\mathfrak{G} = \sum_T \mathfrak{H} T$ is the right co-set decomposition of $\mathfrak{G}$ with respect to $\mathfrak{H}$. Applying this relation to a system of basis elements $h_1, h_2, \cdots, h_m$ of K_1/F and noticing that the discriminant $| h_i^T |^2 \neq 0$, we find $\sum_P a^{PT} \beta_{PT} = 0$, or

$$\sum_P a^P \beta_{PT} = 0 \qquad \left(\sum_S S\beta_S \in \mathfrak{r}\right) \tag{4}$$

for all T (mod $\mathfrak{H}$). Since conversely (4) implies (3), the elements a of $\mathfrak{m}$ can be characterized also by (4). Hence the image of $\mathfrak{m}$ by an isomorphism between K/K_1 and $\mathfrak{H}(K_1)$ consists of all left annihilators of the elements $\sum_P P\beta_{PT}$ $\left(\sum_S S\beta_S \in \mathfrak{r}\right)$, and is a left ideal in $\mathfrak{H}(K_1)$. Hence $\mathfrak{m}$ is a Galois module of K/K_1. The second half of the theorem can be proved also in the same way as in the original paper of Deuring (p. 46).[12]

3. Appendix: On normal bases

Deuring's second proof of the theorem of normal bases, which was published in the same paper Deuring (21) also under the assumption that the characteristic does not divide the degree of extension, can be so modified that it works generally. Furthermore, R. Stauffer's[13] method for constructing a normal basis works, after a modification, also without that same assumption. These will be seen in the following.

Let $\mathfrak{G}$ be, as above, the Galois group of a Galois extension K/F. On taking a sufficiently large over-field F' of F such that all absolutely irreducible representations of $\mathfrak{G}$ lie in F', we consider the algebra $K' = K \times F'$ over F' instead of K itself. It is a right representation module of $\mathfrak{G}$ and we denote by $M(S)$ $(S \in \mathfrak{G})$ the corresponding representation of $\mathfrak{G}$. Let $G(S)$ be an irreducible representation of $\mathfrak{G}$ (in F') with a degree, say g, and let $U(S)$ be the corresponding directly indecomposable part of the regular representation of $\mathfrak{G}$; $U(S)$ has $G(S)$ as its first and its last irreducible as well as largest completely reducible components. We wish to show that $U(S)$ is contained in $M(S)$ (at least) g times.—If this is proved for each irreducible representation $G(S)$, then we find that the regular representation of $\mathfrak{G}$ is contained in, whence coincides

[12] Theorem remains true if we, as in Deuring (21), extend the underlying field F to an over-field F' and consider $K \times F'$, $K_1 \times F'$ instead of K, K_1.

[13] Stauffer (28).

with, $M(S)$ and thus we are through.[14]—That $U(S)$ is contained in $M(S)$ g times is equivalent to that there exists in K' a matrix B with g columns, satisfying

$$B^S = U(S)B \quad \text{for all} \quad S \in \mathfrak{G} \tag{5}$$

and whose elements are all linearly independent (with respect to F'). Now, if we denote by B_1 the matrix consisting of the first g rows of a matrix B satisfying (5), then B_1 fulfills $B_1^S = G(S)B_1$ for all S.[15] Furthermore, since the representation module of $U(S)$ possesses a *unique* simple submodule (corresponding to $G(S)$), it follows easily that in order to show that all the elements of B are linearly independent we have merely to show that the elements of B_1 are linearly independent. And, in order to secure the latter we have in turn only to make sure that all the columns of B_1 are linearly independent (with respect to F'); this is due to the fundamental fact that every automorphism of an (irreducible) representation module of $G(S)$ can be obtained by multiplication with an element of F' (cf. Deuring (21), pp. 43–44 and Stauffer (28), pp. 592–593). Hence our aim is to prove the existence of B with g columns, satisfying (5) and such that the g columns of B_1 are linearly independent with respect to F'.

For this purpose, we can first apply *Speiser's theorem*. (In a *generalized* form the theorem asserts: Let $M_E, M_S, \cdots, M_T$ and $N_E, N_S, \cdots, N_T$ ($\mathfrak{G} = \{E, S, \cdots, T\}$) be two systems of non-singular matrices with elements from K' such that

$$M_S^T M_T M_{ST}^{-1} = N_S^T N_T N_{ST}^{-1}$$

for every pair $S, T \in \mathfrak{G}$. Then there exists a non-singular matrix C which satisfies $C^S M_S C^{-1} = N_S$ for all $S \in \mathfrak{G}$.)[16] Namely, there is a non-singular square matrix C fulfilling $C^S C^{-1} = U(S)$ (that is, $C^S = U(S)C$) for all S. We choose g columns in such a C so that the square matrix C_0 consisting of g^2 elements at the intersections of those g columns with the first g rows is non-singular; this is obviously possible. Denote by B the matrix consisting of the chosen g columns of C. Then this B satisfies the above requirements, because (5) is evidently valid and the corresponding B_1 is simply C_0 whose g columns are linearly independent even with respect to K'.

---

[14] If a representation contains $U(S)$ then $U(S)$ is a direct constituent. See Nakayama-Nesbitt (11), §2.

[15] Here we agree that $U(S)$ has already assumed the form where the right upper part is 0.

[16] This can be proved in exactly the same manner as in the special case Schur (27) (or Weyl (30)), provided that F' contains sufficiently many elements. The case where this last condition is not satisfied can be reduced to a favorable case by an argument due to R. Brauer and E. Noether (cf. Deuring (20) or Van der Waerden (29), p. 70).

The present form, which is more general than needed here and which is also more general than the one given in Deuring (21), for instance, has a significance for the classification of semi-linear transformations; cf. Nakayama (25), Haantjes (22).

A second method to obtain a relevant B is to consider a matrix

$$B = B(z) = \sum_S \tilde{U}(S^{-1})z^S, \qquad z \in K,$$

where by $\tilde{U}(S^{-1})$ we denote the matrix composed of the first g columns of $U(S^{-1})$. This $B = B(z)$ certainly satisfies (5), and the corresponding B_1 is $B_1(z) = \sum_S G(S^{-1})z^S$. Now, $B_1(z)$ satisfies not only the relation $B_1(z)^S = G(S)B_1(z)$ but also the relation

$$B_1(z^S) = B_1(z)G(S).$$

From this fact we can conclude that all the elements of $B_1(z)$ are linearly independent (with respect to F') whenever one of them is different from 0, that is, whenever $B_1(z) \neq 0$. But from the separability of K/F we can see easily that there exists certainly a $z \in K$ such that $B_1(z) \neq 0$. (For all this cf. Stauffer (28)). Thus $B = B(z)$ with such a z possesses the desired properties.

Furthermore, let $G^{(1)}, G^{(2)}, \cdots, G^{(m)}$ be the totality of distinct irreducible representations of $\mathfrak{G}$ in F', and take for each $G(S) = G^{(\mu)}(S)$ an element $z = z^{(\mu)}(\in K)$ satisfying the above condition. Here we can, and shall, choose these $z^{(\mu)}$ in such a manner that $z^{(\mu)} = z^{(\nu)}$ if $G^{(\mu)}$ and $G^{(\nu)}$ are conjugate with respect to F. Let now $G_1, G_2, \cdots, G_k$ be the totality of distinct irreducible representations of $\mathfrak{G}$ in F, and put $z_\kappa = z^{(\mu)}$ if G_κ contains $G^{(\mu)}$. Let further $E_1, E_2, \cdots, E_k$ be mutually orthogonal idempotent elements in the group algebra $\mathfrak{G}(F)$ with the sum $\sum E_\kappa = E = 1$ such that if N denotes the radical of $\mathfrak{G}(F)$ then $E_\kappa \pmod N$ is the unit element in the simple two-sided ideal of $\mathfrak{G}(F)/N$ belonging to $\mathfrak{G}_\kappa$. ($E_\kappa\mathfrak{G}(F)$ is a direct sum of right ideals which define the component U_κ of the regular representation corresponding to G_κ.) Then $w = \sum_\kappa z_\kappa^{E_\kappa}(\in K)$ and its conjugates form a normal basis of K/F. This can be seen in quite a similar way as in Stauffer (28).

Chapter II: Frobeniusean Rings

4. Frobeniusean and quasi-Frobeniusean rings

Consider a ring A. A may have a left, say, operator domain Ω such that $\alpha(a + b) = \alpha a + \alpha b$, $\alpha(ab) = (\alpha a)b = a(\alpha b)$ if $\alpha \in \Omega$ and $a, b \in A$. In case A satisfies both the minimum and the maximum conditions for left and right ideals (allowable with respect to Ω), its general structure is well known. But recently C. Hopkins showed that the theory remains valid to a large extent in the case where A fulfills the minimum condition only.[17] We assume in the present treatment also merely the *minimum condition*. And, since we shall deal exclusively with such type of rings, we shall understand by a *ring* always a one of the type.

A possesses the radical N which is nilpotent and the residue class ring $\bar{A} = A/N$ is semisimple. Everything in Part I, 1 remains valid[18] for our A

[17] Hopkins (23), (24).

[18] With trivial modifications of terminologies, of course.

except the second half of Lemma 1 and the statements concerning the completely reducible ideals. Thus we retain the old significances of the symbols k, $f(\kappa)$, $e_{\kappa,i}$, $e_\kappa(=e_{\kappa,1})$, $E_\kappa = \sum_{i=1}^{f(\kappa)} e_{\kappa,i}$, $E = \sum_{\kappa=1}^{k} E_\kappa$, $c_{\kappa,ij}$ ($\kappa = 1, 2, \cdots, k$; $i, j = 1, 2, \cdots, f(\kappa)$). Namely

$$\bar{A} = \bar{A}_1 + \bar{A}_2 + \cdots + \bar{A}_k,$$

where each $\bar{A}_\kappa$ is simple and has a unit element $\bar{E}_\kappa = E_\kappa \pmod N$; $e_{\kappa,i}$ are mutually orthogonal idempotent elements, and $c_{\kappa,ii} = e_{\kappa,i}$, $c_{\kappa,ij}c_{\lambda,hl} = \delta_{\kappa\lambda}\delta_{jh}c_{\kappa,il}$. As to the connection between $r(N)$ and completely reducible left ideals, we can assert that any completely reducible left ideal is contained in $r(N)$, while $r(N)$ is a direct sum of a completely reducible left ideal and a left ideal which is annihilated by A; this last left ideal is not necessarily completely reducible this time.[19]

Remark. If A possesses a *unit element* then it satisfies also the maximum condition for left and right ideals.[20] Moreover, the same assumption assures of course that $r(N)$ is really the largest completely reducible left ideal.

If $\mathfrak{m}$ is a simple left module of A, then we denote by $d_l(\mathfrak{m})$ the rank of $\mathfrak{m}$ with respect to the quasi-field of automorphisms. Hence, if $\mathfrak{m} \cong \bar{A}\bar{e}_\kappa = Ae_\kappa/Ne_\kappa$ then $d_l(\mathfrak{m}) = f(\kappa)$. More generally, if $\mathfrak{m}$ is a left module which has a composition series $\mathfrak{m} = \mathfrak{m}_0 \supset \mathfrak{m}_1 \supset \cdots \supset \mathfrak{m}_s \supset 0$, then we put $d_l(\mathfrak{m}) = \sum_{i=1}^{s} d_l(\mathfrak{m}_{i-1}/\mathfrak{m}_i)$. For a right module $\mathfrak{n}$ possessing a composition series we define $d_r(\mathfrak{n})$ in the same manner.

Now, since a direct generalization of the definition of Frobeniusean and quasi-Frobeniusean algebras to our general case seems difficult, let us, in view of I, Lemma 2, define the corresponding types of rings as follows:

Definition. A is called *quasi-Frobeniusean* if it possesses a unit element and if there exists a permutation $(\pi(1), \pi(2), \cdots, \pi(k))$ of $(1, 2, \cdots, k)$ such that for each κ

i) $e_\kappa A$ has a unique simple right subideal $\mathfrak{r}_\kappa$ and $\mathfrak{r}_\kappa \cong \bar{e}_{\pi(\kappa)}\bar{A}$,

ii) $Ae_{\pi(\kappa)}$ has a unique simple left subideal $\mathfrak{l}_{\pi(\kappa)}$ and $\mathfrak{l}_{\pi(\kappa)} \cong \bar{A}\bar{e}_\kappa$.

If moreover

iii) $f(\kappa) = f(\pi(\kappa))$,

then we call A *Frobeniusean*.

Remark.[20a] In the above definition we can omit the condition $\mathfrak{l}_{\pi(\kappa)} \cong \bar{A}\bar{e}_\kappa$ in ii). For, if the other conditions are satisfied, that is, if A has a unit element and if i) and the first part of ii) are the case, then $e_\kappa\mathfrak{l}_{\pi(\kappa)} = e_\kappa r(N)e_{\pi(\kappa)} \supseteqq \mathfrak{r}_\kappa e_{\pi(\kappa)} \neq 0$, whence necessarily $\mathfrak{l}_{\pi(\kappa)} \cong \bar{A}\bar{e}_\kappa$.

[19] This is due to our not assuming anything about the structure of Ω.

[20] See Hopkins (24). As a matter of fact, our main interest in the present paper lies in such an A which, either by its definition or as a consequence of its definition, possesses a unit element. Hence our present avoidance of maximum condition is not so important, according to this remark.

[20a] See a correction at the end (Added in proof).

We have now the following theorems which correspond to Th. 1, 2 and 3 in Part I:

THEOREM 6. *If a ring A is quasi-Frobeniusean, then*

$$\alpha) \qquad l(r(\mathfrak{l})) = \mathfrak{l}, \qquad r(l(\mathfrak{r})) = \mathfrak{r}$$

for every left ideal $\mathfrak{l}$ and right ideal $\mathfrak{r}$. Conversely, if α) holds for every nilpotent simple left ideal $\mathfrak{l}$ and nilpotent simple right ideal $\mathfrak{r}$ as well as for $\mathfrak{l} = \mathfrak{r} = N$ (radical of A) and $\mathfrak{l} = \mathfrak{r} = 0$, then A is quasi-Frobeniusean.

THEOREM 7. *If A is Frobeniusean, then besides α) we have*

$$\beta') \qquad d_l(\mathfrak{l}) = d_r(A/r(\mathfrak{l})), \qquad d_r(\mathfrak{r}) = d_l(A/l(\mathfrak{r}))$$

for every left ideal $\mathfrak{l}$ and right ideal $\mathfrak{r}$.[21] *Conversely, if α) is valid for every nilpotent simple left ideal $\mathfrak{l}$, nilpotent simple right ideal $\mathfrak{r}$, $\mathfrak{l} = \mathfrak{r} = N$ and $\mathfrak{l} = \mathfrak{r} = N$ while β') is the case for every nilpotent simple left ideal $\mathfrak{l}$ and nilpotent simple right ideal $\mathfrak{r}$, then A is Frobeniusean.*[22]

PROOF. The second parts of the theorems can be proved in the same way as in I (with a very slight and trivial modification).[23]

To prove the first parts, assume that A is quasi-Frobeniusean. Since $l(N)$ is a two-sided ideal we have $l(N) = \sum_\kappa E_\kappa l(N) = \sum_{\kappa,i} e_{\kappa,i}l(N)$. From the definition of a quasi-Frobeniusean ring it follows that for each κ, i the right ideal $e_{\kappa,i}l(N) = e_{\kappa,i}A \cap l(N)$ is simple and is isomorphic to $\bar{e}_{\pi(\kappa)}\bar{A}$. Hence $e_{\kappa,i}l(N) = e_{\kappa,i}l(N)E_{\pi(\kappa)}$, whereas $e_{\kappa,i}l(N)E_\lambda = 0$ if $\lambda \neq \pi(\kappa)$. Thus $E_\kappa l(N)E_\lambda = E_\kappa l(N)$ or $= 0$ according as $\lambda = \pi(\kappa)$ or not. Since this is the case for every κ, we have $l(N)E_{\pi(\kappa)} = \sum_\mu E_\mu l(N)E_{\pi(\kappa)} = E_\kappa l(N)E_{\pi(\kappa)} = E_\kappa l(N)$. This shows that $E_\kappa l(N)$ is two-sided. Moreover, it is a simple two-sided ideal. To see this, let d be any non-zero element in $E_\kappa l(N)$. $d = \sum_{i=1}^{f(\kappa)} e_{\kappa,i}d$ and at least one of $e_{\kappa,i}d$ is not zero. Suppose $e_{\kappa,p}d \neq 0$. Then $e_{\kappa,p}dA = e_{\kappa,p}l(N)$ since $e_{\kappa,p}l(N)$ is a simple right ideal, and therefore $AdA = Ae_{\kappa,p}l(N) \supseteqq E_\kappa l(N)$. Hence the two-sided ideal $E_\kappa l(N)$ is simple. In particular, it is completely reducible as a left ideal too, that is, $E_\kappa l(N) \subseteq r(N)$. Because this is true for every κ, it follows that $l(N) \subseteq r(N)$. But the inclusion of the other direction can be seen in the same way, and we have $l(N) = r(N)$. We denote this two-sided ideal by M.

Let $\mathfrak{l}$ be any non-zero left ideal of A and let $\mathfrak{l}'$ be a maximal left subideal of $\mathfrak{l}$. Suppose $\mathfrak{l}/\mathfrak{l}' \cong \bar{A}\bar{e}_\kappa$. If $c \in A$ then the left ideal $\mathfrak{l}c$ is isomorphic to $\mathfrak{l}/\mathfrak{l} \cap l(c)$, as can be seen from the mapping $b \to bc$ $(b \in \mathfrak{l})$. If in particular $c \in r(\mathfrak{l}')$ then $\mathfrak{l} \cap l(c) \supseteqq \mathfrak{l} \cap l(r(\mathfrak{l}')) \supseteqq \mathfrak{l}'$ whence $\mathfrak{l}c \cong \bar{A}\bar{e}_\kappa$ or $= 0$. This shows that $\mathfrak{l}r(\mathfrak{l}') \subseteq M$. Let now b_1 be an element in $e_\kappa\mathfrak{l}$ which is not contained in $\mathfrak{l}'$; the existence of such

[21] In particular, we have $d_l(A) = d_r(A)$.

[22] It is justified to speak of d_l and d_r here, since the validity of α) for nilpotent simple ideals as well as for N and 0 implies that A is quasi-Frobeniusean and in particular that A satisfies, because of the existence of a unit element, the maximum condition too.

[23] One has to modify slightly and in an obvious manner the very last part of i) there, while the final step viii) becomes a mere triviality this time.

a b_1 follows from $\mathfrak{l}/\mathfrak{l}' \cong \bar{A}\bar{e}_\kappa$. Then $Ab_1 \cup \mathfrak{l}' = \mathfrak{l}$. From the above consideration we have $b_1 r(\mathfrak{l}') \subseteq e_\kappa \mathfrak{l} r(\mathfrak{l}') \subseteq e_\kappa M$ whence $b_1 r(\mathfrak{l}') = e_\kappa M$ or $= 0$. Thus $b_1 r(\mathfrak{l}') \cong \bar{e}_{\pi(\kappa)}\bar{A}$ or $= 0$. On the other hand $b_1 r(\mathfrak{l}') \cong r(\mathfrak{l}')/r(\mathfrak{l}') \cap r(b_1)$ and here $r(\mathfrak{l}') \cap r(b_1) = r(\mathfrak{l}') \cap r(Ab_1) = r(\mathfrak{l}' \cup Ab_1) = r(\mathfrak{l})$. Hence finally $r(\mathfrak{l}')/r(\mathfrak{l}) \cong \bar{e}_{\pi(\kappa)}\bar{A}$ or $= 0$.

Consider a composition series of left ideals $\mathfrak{l}_0 = 0 \subset \mathfrak{l}_1 \subset \mathfrak{l}_2 \subset \cdots \subset \mathfrak{l}_s = A$ of A. The above observation shows that each right module $r(\mathfrak{l}_t)/r(\mathfrak{l}_{t+1})$ is either simple or 0. Since $r(0) = A$ and $r(A) = 0$, it follows that the length of a composition series of right ideals of A is less than or equal to s, the length of a composition series of left ideals. But the inequality of the other direction can be seen in the same manner. Hence the lengths of composition series of right and left ideals are equal to each other, and moreover the right moduli $r(\mathfrak{l}_t)/r(\mathfrak{l}_{t+1})$ $(t = 0, 1, \cdots, s-1)$ are all simple. Further, all the left moduli $l(r(\mathfrak{l}_{t+1}))/l(r(\mathfrak{l}_t))$ must be also simple, and necessarily $l(r(\mathfrak{l}_t)) = \mathfrak{l}$ for all t. Since there exists always at least on composition series through any given left ideal, we have $l(r(\mathfrak{l})) = \mathfrak{l}$ for any left ideal $\mathfrak{l}$. Similarly $r(l(\mathfrak{r})) = \mathfrak{r}$ for every right ideal $\mathfrak{r}$. That is, α) is always valid in A. This proves the first part of Theorem 6.

Our proof shows further that if $\mathfrak{l}/\mathfrak{l}' \cong \bar{A}\bar{e}_\kappa$ then $r(\mathfrak{l}')/r(\mathfrak{l}) \cong \bar{e}_{\pi(\kappa)}\bar{A}$. And this fact, together with the relation $d_l(\bar{A}\bar{e}_\kappa) = d_r(\bar{e}_\kappa\bar{A}) = f(\kappa)$, shows the validity of β') in case of a (not only quasi-Frobeniusean but) Frobeniusean ring A.

5. Corollaries

THEOREM 8. *Let A be a quasi-Frobeniusean ring. The composition length of a principal left ideal Ac is equal to that of the principal right ideal cA. If $c = c_1 + c_2$ and Ac is the direct sum $Ac = Ac_1 + Ac_2$, then cA is the direct sum $cA = c_1A + c_2A$. Further, if A is not only quasi-Frobeniusean but Frobeniusean, then $d_l(Ac) = d_r(cA)$.*

PROOF. Cf. I, Th. 4.

THEOREM 9. *Let A be quasi-Frobeniusean, and let $\pi(\kappa)$ have the same significance as in the definition. If $\mathfrak{l}'$ is a maximal left subideal of a left ideal $\mathfrak{l}$ in A and if the left module $\mathfrak{l}/\mathfrak{l}'$ is isomorphic to $\bar{A}\bar{e}_\kappa$, then the right module $r(\mathfrak{l}')/r(\mathfrak{l})$ is isomorphic to $\bar{e}_{\pi(\kappa)}\bar{A}$. The quasi-field of automorphisms of $\mathfrak{l}/\mathfrak{l}'$ is inversely isomorphic to that of $r(\mathfrak{l}')/r(\mathfrak{l})$. Furthermore, $r(N^\nu) = l(N^\nu)$ for any $\nu = 1, 2, \cdots$; where N is, as before, the radical of A.*

PROOF. The first assertion was shown in the proof of Th. 6. The second one is, by virtue of the first one, equivalent to the relation $\bar{e}_\kappa\bar{A}\bar{e}_\kappa \cong \bar{e}_{\pi(\kappa)}\bar{A}\bar{e}_{\pi(\kappa)}$, and this in turn can be seen in the same way as in I, Th. 3, vii). The final statement follows from the special case $r(N) = l(N)$, which was already established in the proof of Th. 6; See I, Th. 6.

THEOREM 10. *If in a ring A the relation α) holds for $\mathfrak{l} = \mathfrak{r} = N$ (radical), $\mathfrak{l} = \mathfrak{r} = 0$, and both α) and β') hold for all nilpotent simple two-sided ideals $\mathfrak{l} = \mathfrak{r} = \mathfrak{z}$, then A is Frobeniusean.*

PROOF. Arguments of i), ii), iii) and iv) of I, Th. 7 can be transferred term

by term. Now, $E_\kappa M(= E_\kappa M E_{\pi(\kappa)}) = \sum_{i=1}^{f(\kappa)} e_{\kappa,i}M$, and here the $f(\kappa)$ right ideals $e_{\kappa,i}M$ are mutually isomorphic and are direct sums of simple right ideals isomorphic to $\bar{e}_{\pi(\kappa)}\bar{A}$. Hence $d_r(E_\kappa M) \geqq f(\kappa)d_r(\bar{e}_{\pi(\kappa)}\bar{A}) = f(\kappa)f(\pi(\kappa))$ and the equality sign holds if and only if $e_{\kappa,i}M$ are simple. But $d_r(E_\kappa M) = d_l(A/l(E_\kappa M)) = d_l(A/N \cup A(E - E_\kappa)A)) = f(\kappa)^2$ by our assumption.[24] Thus $f(\kappa) \geqq f(\pi(\kappa))$. Since this is the case for every κ, necessarily $f(\kappa) = f(\pi(\kappa))$ and $e_{\kappa,i}M$ are simple. Similarly we find that $Me_{\kappa,i}$ are simple left ideals. It follows then easily that A is Frobeniusean.

THEOREM 11. *Suppose that* α) *holds for every nilpotent simple two-sided ideal* $\mathfrak{l} = \mathfrak{r} = \mathfrak{z}$ *as well as for* $\mathfrak{l} = \mathfrak{r} = N$, $\mathfrak{l} = \mathfrak{r} = 0$. *Let moreover*

$$\gamma) \qquad d_r(\mathfrak{z}) = d_r(A/r(\mathfrak{z})), \qquad d_l(\mathfrak{z}) = d_l(A/l(\mathfrak{z}))$$

for every nilpotent simple two-sided ideal $\mathfrak{z}$. *Then the ring* A *is Frobeniusean. Furthermore, the same is true when we replace the above* γ) *by*

$$\gamma') \qquad d_l(\mathfrak{z}) = d_l(A/r(\mathfrak{z})), \qquad d_r(\mathfrak{z}) = d_r(A/l(\mathfrak{z}))$$

or by

$$\gamma'') \qquad d_r(\mathfrak{z}) = d_l(A/r(\mathfrak{z})), \qquad d_l(\mathfrak{z}) = d_r(A/l(\mathfrak{z})).$$

PROOF. Arguments in i), ii), $\cdots$, iv) of I, Th. 7 remain again valid. And, again $d_r(E_\kappa M) \geqq f(\kappa)d_r(\bar{e}_{\pi(\kappa)}\bar{A}) = f(\kappa)f(\pi(\kappa))$; the equality sign is true if and only if $e_{\kappa,i}M$ are simple. Moreover, $d_r(A/r(E_\kappa M)) = d_r(A/N \cup A(E - E_{\pi(\kappa)})A) = f(\pi(\kappa))^e$, $d_r(A/l(E_\kappa M)) = d_r(A/N \cup A(E - E_\kappa)A) = f(\kappa)^2$ and $d_l(A/r(E_\kappa M)) = f(\pi(\kappa))^2$. Our assertions follow from these relations in quite a similar manner as above.

6. Vector moduli and a theorem of M. Hall

Let g be a natural number and consider a module $\mathfrak{A}$ consisting of all g-dimensional vectors $\mathfrak{x} = (x_1, x_2, \cdots, x_g)$ with components x_p from A. For a subset $\mathfrak{S}$ of A we denote by $R(\mathfrak{S})[L(\mathfrak{S})]$ the set of vectors $\mathfrak{y} = (y_1, y_2, \cdots, y_g)$ such that (the scalar product) $(\mathfrak{x}, \mathfrak{y}) = \sum_{p=1}^{g} x_p y_p[(\mathfrak{y}, \mathfrak{x}) = \sum_p y_p x_p] = 0$ for every $\mathfrak{x} \in \mathfrak{S}$. If we take our theorem 6 into account, the first part of the following theorem is equivalent to the principal theorem (Th. 5. 2) in Hall (7):

THEOREM 12. *Let* A *be quasi-Frobeniusean. Then* $L(R(\mathfrak{L})) = \mathfrak{L}$, $R(L(\mathfrak{R})) = \mathfrak{R}$ *for every* A*-left-submodule* $\mathfrak{L}$ *and* A*-right-submodule* $\mathfrak{R}$ *in* $\mathfrak{A}$. *If moreover* A *is Frobeniusean then* $d_l(\mathfrak{L}) = d_r(\mathfrak{A}/R(\mathfrak{L}))$ *and* $d_r(\mathfrak{R}) = d_l(\mathfrak{A}/L(\mathfrak{R}))$.

We want to show that we can derive this theorem also from our main theorems 6, 7, obtaining thus a second proof, though rather long, of Hall's theorem. For that purpose, consider the g-rowed matrix ring $B = A_g = \sum_{p,q=1}^{g} \epsilon_{pq}A$ over the quasi-Frobeniusean ring A; where ϵ_{pq} is a system of matrix units commutative with every element of A.

LEMMA 2. *B is quasi-Frobeniusean.*

[24] Cf. I, Th. 7, ii).

PROOF. Evidently $\epsilon_{pp}e_{\kappa,i}$ $(p = 1, 2, \cdots, g; \kappa = 1, 2, \cdots, k; i = 1, 2, \cdots, f(\kappa))$ form a system of mutually orthogonal idempotent elements whose sum is the unit element of B. $Q = \sum \epsilon_{pq}N$ is the radical of B and $r(\sum \epsilon_{pq}N) = l(\sum \epsilon_{pq}N) = \sum \epsilon_{pq}M$; where N is, as before, the radical of A and M is the two-sided ideal $r(N) = l(N)$ in A. We put $P = \sum \epsilon_{pq}M$. From $e_\kappa M \cong e_{\pi(\kappa)}A/e_{\pi(\kappa)}N$ (with respect to A), it follows easily that $\epsilon_{pp}e_\kappa P \cong \epsilon_{pp}e_{\pi(\kappa)}B/\epsilon_{pp}e_{\pi(\kappa)}Q \cong \epsilon_{11}e_{\pi(\kappa)}B/\epsilon_{11}e_{\pi(\kappa)}Q$ with respect to B. Similarly $P\epsilon_{pp}e_{\pi(\kappa)} \cong B\epsilon_{11}e_\kappa/Q\epsilon_{11}e_\kappa$. But this shows that B is quasi-Frobeniusean.

Now, consider the A-left-module $\mathfrak{A}_1 = \epsilon_{11}B = \sum_p \epsilon_{1p}A$ and the A-right-module $\mathfrak{A}_2 = B\epsilon_{11} = \sum_p \epsilon_{p1}A$. The mapping θ_1: $\mathfrak{x} = (x_p) \rightarrow \sum \epsilon_{1p}x_p$ is an operator isomorphism of $\mathfrak{A}$ and $\mathfrak{A}_1$ with respect to the left operator ring A. Similarly, $\theta_2 : \mathfrak{y} = (y_p) \rightarrow \sum \epsilon_{1p}y_p$ is an operator isomorphism of $\mathfrak{A}$ onto $\mathfrak{A}_2$ with respect to the right operator ring A. Moreover, $\theta_1(\mathfrak{x})\theta_2(\mathfrak{y}) = \sum \epsilon_{1p}x_py_p\epsilon_{p1} = (\mathfrak{x}, \mathfrak{y})\epsilon_{11}$. In particular $\theta_1(\mathfrak{x})\theta_2(\mathfrak{y}) = 0$ if and only if $(\mathfrak{x}, \mathfrak{y}) = 0$. Hence $\theta_2(R(\mathfrak{S}))(\mathfrak{S} \subseteq \mathfrak{A})$ is the set of right annihilators of $\theta_1(\mathfrak{S})$ in $\mathfrak{A}_2$, in the sense of multiplication in B. That is,

$$\theta_2(R(\mathfrak{S})) = \mathfrak{A}_2 \cap r(\theta_1(\mathfrak{S})). \tag{6}$$

Similarly

$$\theta_1(L(\mathfrak{S})) = \mathfrak{A}_1 \cap l(\theta_2(\mathfrak{S})). \tag{7}$$

On the other hand, $B(\mathfrak{l} \cap \mathfrak{A}_1) = \mathfrak{l}$ for every left ideal $\mathfrak{l}$ in B, while $B\mathfrak{L}_1 \cap \mathfrak{A}_1 = \mathfrak{L}_1$ for every (A-left-) submodule $\mathfrak{L}_1$ of $\mathfrak{A}$. (Because $\mathfrak{l} \cap \mathfrak{A}_1 = \epsilon_{11}\mathfrak{l}$ whence $B(\mathfrak{l} \cap \mathfrak{A}_1) = B\epsilon_{11}\mathfrak{l} = B\epsilon_{11}B\mathfrak{l} = B\mathfrak{l} = \mathfrak{l}$, and $B\mathfrak{L}_1 \cap \mathfrak{A}_1 = \epsilon_{11}B\mathfrak{L}_1 = \epsilon_{11}B\epsilon_{11}\mathfrak{L}_1 = A\epsilon_{11}\mathfrak{L}_1 = A\mathfrak{L}_1 = \mathfrak{L}_1$.) Similarly $(\mathfrak{r} \cap \mathfrak{A}_2)B = \mathfrak{r}$ and $\mathfrak{R}_2B \cap \mathfrak{A}_2 = \mathfrak{R}_2$ for every right ideal $\mathfrak{r}$ in B and (A-right-) submodule $\mathfrak{R}_2$ in $\mathfrak{A}_2$. We have now

$$\begin{aligned}
\theta_1(L(R(\mathfrak{L}))) &= \mathfrak{A}_1 \cap l(\theta_2(R(\mathfrak{L}))) && \text{by (7)}\\
&= \mathfrak{A}_1 \cap l(\mathfrak{A}_2 \cap r(\theta_1(\mathfrak{L}))) && \text{by (6)}\\
&= \mathfrak{A}_1 \cap l((\mathfrak{A}_2 \cap r(B\theta_1(\mathfrak{L})))B) = \mathfrak{A}_1 \cap l(r(B\theta_1(\mathfrak{L})))\\
&= \mathfrak{A}_1 \cap B\theta_1(\mathfrak{L}) && \text{(since } B \text{ is } \textit{quasi-Frobeniusean}\text{)}\\
&= \theta_1(\mathfrak{L}),
\end{aligned}$$

whence $L(R(\mathfrak{L})) = \mathfrak{L}$. Similarly $R(L(\mathfrak{R})) = \mathfrak{R}$, and this proves the first part of the theorem. To prove the second part, one has only to notice that if A is Frobeniusean then B is so too, as can readily be seen, and that $d_l(B\mathfrak{L}_1)$ with respect to B is equal to $d_l(\mathfrak{L}_1)g$ with respect to A.

COROLLARY.[25] *Let $\mathfrak{G}$ be a finite group and let $\mathfrak{G}(A)$ be the group ring of $\mathfrak{G}$ over A. If A is quasi-Frobeniusean then $\mathfrak{G}(A)$ is so too.*

[25] This corollary, as well as the above lemma 2, is closely related to the next section.

PROOF. Let $\mathfrak{G} = \{G_1, G_2, \cdots, G_g\}$. Let $\mathfrak{L}$ be a left ideal in $\mathfrak{G}(A)$. We want to show that an element $Y = \sum_p y_p G_p^{-1}$ in $\mathfrak{G}(A)$ is a right annihilator of $\mathfrak{L}$ if and only if $\sum x_p y_p = 0$ for all $X = \sum x_p G_p$ in $\mathfrak{L}$; if we succeed in showing this and the corresponding fact for a right ideal, then our corollary is an immediate consequence of Th. 6 and 12. Now, since $\sum x_p y_p$ is the coefficient of the group unit element in the product XY, the "only if" part is trivial. To prove the "if" part, suppose that Y satisfies the above condition. Let X be an arbitrary element in $\mathfrak{L}$. Then $G_p^{-1}X \in \mathfrak{L}$ for all $p = 1, 2, \cdots, g$. But the coefficient of G_p in XY is equal to the coefficient of the unit element in $G_p^{-1}XY$. It follows then that all the coefficients in XY vanish, that is, $XY = 0$. Hence $\mathfrak{L}Y = 0$.

Supplement for the case of an algebra. If A is a *Frobeniusean algebra* over a field F, then we can evidently replace d_l and d_r in the above theorem 12 by the dimension with respect to F. Furthermore, in this case we have the following generalization of Th. 2:

THEOREM 13. *Let A be a Frobeniusean algebra. Let $\mathfrak{L}$, $\mathfrak{L}_0$ be A-left-submoduli of the vector module $\mathfrak{A}$ such that $\mathfrak{L} \supseteqq \mathfrak{L}_0$, and denote the representations of A defined by the left module $\mathfrak{L}/\mathfrak{L}_0$ and the right module $R(\mathfrak{L}_0)/R(\mathfrak{L})$ by $a \to M(a)$ and $a \to N(a)$. Then the first representation $a \to M(a)$ is equivalent to the representation $a^* \to N(a)$, where $a \to a^*$ is the automorphism of A given in Theorem 1.*

PROOF. On retaining the notations of the section 1, we consider the linear function $\lambda((\mathfrak{x}, \mathfrak{y})) = \lambda(\sum x_p y_p)$ of the scalar product $(\mathfrak{x}, \mathfrak{y})$. And, for a subset $\mathfrak{S}$ in $\mathfrak{A}$ we denote by $\mathfrak{p}(\mathfrak{S})[\mathfrak{q}(\mathfrak{S})]$ the set of $\mathfrak{y}$ such that $\lambda((\mathfrak{x}, \mathfrak{y})) = 0$ $[\lambda((\mathfrak{y}, \mathfrak{x})) = 0]$ for all $\mathfrak{x} \in \mathfrak{S}$. It follows in quite a similar manner as before[26] that $\mathfrak{q}(\mathfrak{L}) = R(\mathfrak{L})$. And, the representation of A defined by the left module $\mathfrak{L}/\mathfrak{L}_0$ is equivalent to the one defined by the right module $\mathfrak{p}(\mathfrak{L}_0)/\mathfrak{p}(\mathfrak{L})$ (Cf. I, 8). Furthermore, φ: $\mathfrak{x} = (x_p) \to \mathfrak{x}^* = (x_p^*) = (\varphi(x_p))$ is a 1-1 mapping of $\mathfrak{A}$ on itself (See Th. 1), and it is characterized also by the relation $\lambda((\mathfrak{x}^*, \mathfrak{y})) = \lambda((\mathfrak{y}, \mathfrak{x}))$ (for all $\mathfrak{y}$). Evidently $\mathfrak{q}(\mathfrak{S})^* = \mathfrak{p}(\mathfrak{S})$, whence $R(\mathfrak{L})^* = \mathfrak{p}(\mathfrak{L})$. Thus the representation of A defined by $\mathfrak{L}/\mathfrak{L}_0$ is equivalent to the one defined by $R(\mathfrak{L}_0)^*/R(\mathfrak{L})^*$, and this is our assertion.

7. Product rings

Suppose that our operator domain Ω of A contains a field[27] F and that every operator in F commutes with every operator in Ω. Consider on the other hand a (finite linear) algebra B over F. By replacing in the usual manner the coefficient domain F by A we obtain a ring $C = B \times A$, which has Ω as a domain of operators.[28] (Namely, if $b_1, b_2, \cdots, b_n$ $(n = (B:F))$ is a basis of B, then

[26] Cf. I, 3.

[27] We naturally suppose that the addition and the multiplication in F coincide with those as operators and that the unit element in F is the identity operator.

[28] The conditions about the operator domain at the beginning of the section 4 are satisfied with regard to Ω and C. Moreover, C fulfills the minimum condition for right and left ideals.

$C = b_1A + b_2A + \cdots + b_nA$. To introduce the law of multiplication in C, we first define the product ba of $b = \sum \xi_i b_i \in B$, $a \in A$ to be $\sum_i b_i(\xi_i a)$ and then define $(\sum b_i a_i)(\sum b_i a_i')$ to be $\sum_{i,j}(b_i b_j)(a_i a_j')$. Further, $\alpha \sum b_i a_i = \sum b_i(\alpha a_i)$ for $\alpha \in \Omega$.)[29]

THEOREM 14. *The ring $C = B \times A$ is quasi-Frobeniusean [Frobeniusean] if and only if both A and B are so.*

PROOF. If $\mathfrak{l}_1$ and $\mathfrak{l}$ are left ideals in B and A respectively, the module product[30] $\mathfrak{l}_1\mathfrak{l}$ is evidently a left ideal in C, and

$$r(\mathfrak{l}_1\mathfrak{l}) = Br(\mathfrak{l}) \cup r(\mathfrak{l}_1)A \tag{8}$$

as can readily be seen.[31] (For a proof, take a basis $b_1, b_2, \cdots, b_n$ of B such that $r(\mathfrak{l}_1)$ is spanned by $b_{s+1}, b_{s+2}, \cdots, b_n$ $(s = n - (r(\mathfrak{l}_1):F))$. Suppose that an element $c = \sum b_i a_i$ in C is a right annihilator of the left ideal $\mathfrak{l}_1\mathfrak{l}$. Namely, if $b \in \mathfrak{l}_1$, $a \in \mathfrak{l}$ then $ba(\sum b_i a_i) = \sum (bb_i)(aa_i) = 0$. Let $bb_i = \sum_j \beta_{ij}(b)b_j$. Then $\sum_{i,j} b_j(\beta_{ij}(b)aa_i) = 0$, or

$$\sum_i \beta_{ij}(b)aa_i = 0 \qquad \text{for all } j = 1, 2, \cdots, n. \tag{9}$$

Here we notice that $\beta_{ij}(b) = 0$ if $i = s+1, s+2, \cdots, n$. Now, let b run over a system of generators $b^{(1)}, b^{(2)}, \cdots, b^{(t)}$ of $\mathfrak{l}_1$. Then the s matrices $B_i = (\beta_{ij}(b^{(p)}))_{jp}$ $(i = 1, 2, \cdots, s)$ are linearly independent, because otherwise a non-trivial linear combination of $b_1, b_2, \cdots, b_s$ would be a right annihilator of $\mathfrak{l}_1$. In other words, the system of linear equations $\sum_{i=1}^{s} x_i\beta_{ij}(b^{(p)}) = 0$ $(j = 1, 2, \cdots, n;\ p = 1, 2, \cdots, t)$ has no non-trivial solution, that is, the rank of the matrix $(\beta_{ij}(b^{(p)}))_{i,jp}$ $(i = 1, 2, \cdots, s)$ is s. But then $\sum_{i=1}^{s} \beta_{ij}(b^{(p)})aa_i = 0$ (cf. (9))[32] implies $aa_i = 0$ $(i = 1, 2, \cdots, s)$. Since a is any element in $\mathfrak{l}$, this shows that $a_1, a_2, \cdots, a_s \in r(\mathfrak{l})$ and $c = \sum b_i a_i = \sum_{i=1}^{s} b_i a_i + \sum_{i=s+1}^{n} b_i a_i \in Br(\mathfrak{l}) \cup r(\mathfrak{l}_1)A$. Thus $r(\mathfrak{l}_1\mathfrak{l}) \subseteq Br(\mathfrak{l}) \cup r(\mathfrak{l}_1)A$. The inclusion of the other direction is trivial.)

a) Now let A and B be quasi-Frobeniusean. Let N, Q be the radical of A, B respectively, and put $M = r(N) = l(N)$, $P = r(Q) = l(Q)$. The two-sided ideal $BN \cup QA$ in $C = B \times A$ is evidently contained in the radical S of C and we have, according to (8), $r(BN \cup QA) = r(BN) \cap r(QA) = BM \cap PA = PM$. Hence $r(S) \subseteq PM$.

Let g and e be any primitive idempotent elements in B and A respectively. There exist, since B and A are quasi-Frobeniusean, primitive idempotent elements g', e' in B, A such that $Pg \cong Bg'/Qg'$ $(\cong Bg' \cup Q/Q)$ with respect to

[29] The structure of C thus obtained is of course independent of the special choice of the basis $b_1, b_2, \cdots, b_n$.

[30] The product of an element in B with an element in A is defined as in the above elucidation of $C = B \times A$.

[31] Here, as well as in the whole following proof, $r(*)$ denotes the set of right annihilators of the argument in A, B or C according as the argument lies in A, B or C.

[32] Cf. also a remark following (9).

(the left operator ring) B and $Me \cong Ae'/Ne'$ ($\cong Ae' \cup N/N$) with respect to A. Then, as can readily be seen,

$$PMge \cong Cg'e'/(BN \cup QA)g'e'(\cong Cg'e' \cup BN \cup QA/BN \cup QA) \tag{10}$$

with respect to C. Now the following fact is more or less well known and is easy to see:

LEMMA 3. *If in particular B and A are semisimple, then $B \times A$ is Frobeniusean; (it is in fact uni-serial (= einreihig) in the sense of Köthe* (9)[33]).

On applying the lemma to the simple algebra $Bg'B \cup Q/Q$ and the simple ring $Ae'A \cup N/N$ (instead of B, A), we find that $Cg'e'/Sg'e'$ is isomorphic, with respect to the left operator ring C, to the largest completely reducible submodule of $Cg'e'/(BN \cup QA)g'e'$. But this latter is, because of (10) and $r(S) \subsetneq PM$, isomorphic to $r(S)ge$, and therefore

$$Cg'e'/Sg'e' \cong r(S)ge. \tag{11}$$

Let $ge = j_1 + j_2 + \cdots + j_p$ and $g'e' = j_1' + j_2' + \cdots + j_q'$ are decompositions of ge and $g'e'$ into mutually orthogonal idempotent elements in C. Then the left side of (11) is a direct sum of exactly q simple submoduli, while the right side is a direct sum of at least p simple submoduli.[34] Hence $p \leqq q$. On the other hand, the right moduli $g'P$, $e'M$ of B, A are isomorphic to gB/gQ, eA/eN respectively and it follows in the same way as above that $q \leqq p$. Thus $q = p$, and all the left ideals $r(S)j_1$, $r(S)j_2$, $\cdots$, $r(S)j_p$ are necessarily simple. Moreover, they are isomorphic, up to their arrangement, with Cj_1'/Sj_1', Cj_2'/Sj_2', $\cdots$, Cj_q'/Sj_q'. Similarly the right ideals $j_1'l(S)$, $j_2'l(S)$, $\cdots$ are simple and isomorphic to j_1C/j_1S, j_2C/j_2S, $\cdots$ except their arrangement.

If we apply this consideration to all pairs of primitive idempotent elements g and e appearing in a decomposition of the unit elements in B and A, then we find readily that $C = B \times A$ is quasi-Frobeniusean, and that it is even Frobeniusean if both B and A are so.

(A second proof of the quasi-Frobeniusean part can be obtained in the following way: First let B be not only quasi-Frobeniusean but Frobeniusean. Then the parastrophic determinant of B is not identically zero, and an easy modification of the proof of the corollary in the preceeding section 6 shows that the relation α) holds in $C = B \times A$, whence C is quasi-Frobeniusean. The case of a quasi-Frobeniusean B can be reduced to this Frobeniusean case by a device similar to that of Part I, 3.)

b) Assume conversely that $C = B \times A$ is quasi-Frobeniusean. Then $r(C) = 0$, whence $r(B) = 0$ and $r(A) = 0$ because of (8). Hence the same relation (8) shows that $r(B\mathfrak{l}) = Br(\mathfrak{l})$ and $r(\mathfrak{l}_1A) = r(\mathfrak{l}_1)A$. If we combine these relations and their (left, right) duals with the relation α) in C, which is valid by our

[33] Cf. also the section 9 below.

[34] Observe that $r(S)j = r(S) \cap Cj$ is the largest completely reducible left subideal of Cj, which is not 0 evidently.

assumption, then we find that α) is valid also in both A and B. Hence A and B are quasi-Frobeniusean.

That if C is Frobeniusean then A and B are so can be seen from the analysis in the above a) which is, since A and B are quasi-Frobeniusean at least, now applicable.

(It is also possible to prove the first part of this b) by a structural analysis similar to a). On the other hand, in case (not only B but) A is an algebra over F, the second part of b) may be seen also by using (8) and the relation β) (I, Th. 1) or β') in C.)

We now prove the following supplement of Th. 12, Corollary:

COROLLARY. *Let A be Frobeniusean. Then the group ring $\mathfrak{G}(A)$ is Frobeniusean too.*

PROOF. The A-left-moduli A/N and $M = r(N) = l(N)$ are isomorphic to each other. It follows easily that the $\mathfrak{G}(A)$-left-moduli $\mathfrak{G}(A/N)$ $(\cong \mathfrak{G}(A)/\mathfrak{G}(N))$ and $\mathfrak{G}(M)$ are isomorphic. In particular, the largest completely reducible $\mathfrak{G}(A)$-left-submodule of $\mathfrak{G}(M)$ is isomorphic to that of $\mathfrak{G}(A/N)$. But the latter is in turn isomorphic to the residue class module of $\mathfrak{G}(A/N)$ with respect to its radical, because the group ring $\mathfrak{G}(A/N)$ over the semisimple ring, A/N is Frobeniusean according to the above theorem 14.[35] Since $\mathfrak{G}(N)$ is evidently contained in the radical S of $\mathfrak{G}(A)$ and since $\mathfrak{G}(M)$ contains the right annihilator $r(S)$ of S in $\mathfrak{G}(A)$, we find that the residue class module $\mathfrak{G}(A)/S$ is isomorphic to the largest completely reducible left ideal of $\mathfrak{G}(A)$, with respect to the left operator domain $\mathfrak{G}(A)$. But we know already that $\mathfrak{G}(A)$ is quasi-Frobeniusean at least. And, these two facts assure that $\mathfrak{G}(A)$ is Frobeniusean.

8. Residue class rings

THEOREM 15. *Let A be a Frobeniusean ring and let $\mathfrak{z}$ be a two sided ideal in A. The residue class ring $\tilde{A} = A/\mathfrak{z}$ is Frobeniusean if and only if the two-sided ideals $r(\mathfrak{z})$ and $l(\mathfrak{z})$ are respectively a principal right and a principal left ideal: $r(\mathfrak{z}) = bA$, $l(\mathfrak{z}) = Ac$.*

PROOF. Let $\{\rho\}$ be the set of such indices that $E_\rho \notin \mathfrak{z}$. Since E_ρ are idempotent, $E_\rho \notin \mathfrak{z} \cup N$ either. (For, $(z + n)^r$ $(z \in \mathfrak{z}, n \in N)$ is a sum of n^r with an element in $\mathfrak{z}$, and n^r vanishes for a sufficiently high r.) Denote the residue classes of E_ρ, $e_{\rho,i}$ (mod $\mathfrak{z}$) by $\tilde{E}_\rho$, $\tilde{e}_{\rho,i}$. $\tilde{N} = N \cup \mathfrak{z}/\mathfrak{z}$ is the radical of $\tilde{A}$, and $\tilde{A}/\tilde{N} = \sum_\rho(\tilde{A}/\tilde{N})E_\rho$ is the (unique) decomposition of $\tilde{A}/\tilde{N}$ into a direct sum of simple two-sided ideals. $\tilde{e}_{\rho,i}$ are primitive idempotent elements in $\tilde{A}$, since $\tilde{A}\tilde{e}_{\rho,i}/N\tilde{e}_{\rho,i}$ are simple. Moreover, $E_\kappa(r(\mathfrak{z}) \cap M) \neq 0$ if and only if $\kappa \in \{\rho\}$, where $M = r(N) = l(N)$ as before, because $l(r(\mathfrak{z}) \cap M) = l(r(\mathfrak{z} \cup N)) = \mathfrak{z} \cup N$.

[35] We decompose first A/N into a direct sum of simple rings and then consider the group rings over those simple components. They are Frobeniusean, because they can be considered as products of those simple rings with the group algebras constructed over their centers, for instance.

On the other hand, M is a direct sum of simple two-sided ideals $E_\kappa M = ME_{\pi(\kappa)}$ and every two-sided subideal of M is, as can readily be seen, a direct sum of some $E_\kappa M$. Hence

$$r(\mathfrak{z}) \cap M = \sum_\rho E_\rho M (= \sum_\rho ME_{\pi(\rho)} = \sum_{\rho,i} Me_{\pi(\rho),i}). \tag{12}$$

a) Assume now $r(\mathfrak{z}) = bA$. We observe first that $e_{\kappa,i}b \neq 0$ if and only if $\kappa \in \{\rho\}$. The left ideal $Ae_{\rho,i}b$ is isomorphic to $A/l(e_{\rho,i}b)$ and here $l(e_{\rho,i}b) = l(e_{\rho,i}bA) = l(e_{\rho,i}A \cap bA)$[36] $= l(e_{\rho,i}A) \cup l(bA) = A(E - e_{\rho,i}) \cup \mathfrak{z}$. That is, $Ae_{\rho,i}b \cong A/A(E - e_{\rho,i}) \cup \mathfrak{z} \cong \bar{A}\bar{e}_{\rho,i}$. Moreover $Ab \cong A/l(b) = A/\mathfrak{z} = \bar{A}$. Since Ab is the sum of $Ae_{\rho,i}b$ while $\bar{A}$ is the direct sum of $\bar{A}\bar{e}_{\rho,i}$, we find that Ab is indeed the direct sum of $Ae_{\rho,i}b$.[37] Each $Ae_{\rho,i}b$ contains at least one simple left subideal, and therefore the largest completely reducible left subideal $Ab \cap M$ of Ab is a direct sum of at least $\sum_\rho f(\rho)$ simple left ideals. But $Ab \cap M \subseteq r(\mathfrak{z}) \cap M$, since $Ab \subseteq r(\mathfrak{z})$, and here $r(\mathfrak{z}) \cap M$ is the direct sum of $\sum_\rho f(\pi(\rho)) = \sum_\rho f(\rho)$ simple left ideals $Me_{\pi(\rho),i}$ (See (12)). Hence necessarily $Ab \cap M = r(\mathfrak{z}) \cap M$ and each $Ae_{\rho,i}b$ has only one simple left subideal. Furthermore, since $Ae_{\rho,1}b \cong \cdots \cong Ae_{\rho,f(\rho)}b$ (for $\bar{A}\bar{e}_{\rho,1} \cong \cdots \cong Ae_{\rho,f(\rho)}$), it follows that there exists a permutation $\{\nu(\rho)\}$ of $\{\rho\}$ such that the unique simple left subideal of $Ae_{\nu(\rho),i}b$ is, for each ρ, i, isomorphic to $Me_{\pi(\rho),i} \cong \bar{A}\bar{e}_\rho \cong \bar{A}\bar{e}_\rho/\bar{N}\bar{e}_\rho$. Thus $\bar{A}\bar{e}_{\nu(\rho),i}$ has also a unique simple left ideal, which is isomorphic to $\bar{A}\bar{e}_\rho/\bar{N}\bar{e}_\rho$. We observe also that $f(\nu(\rho)) = f(\pi(\rho)) = f(\rho)$ according to our construction.

Assume further $l(\mathfrak{z}) = Ac$. Then we find in the same way as above that every $\bar{e}_\rho\bar{A}$ has only one simple right subideal, and this simple right ideal is, by a remark in 4, necessarily isomorphic to $\bar{e}_{\nu(\rho)}\bar{A}/\bar{e}_{\nu(\rho)}\bar{N}$. $\bar{A}$ is therefore Frobeniusean.

b) To prove the converse, suppose that $\bar{A} = A/\mathfrak{z}$ is Frobeniusean. Let $\bar{P}$ be the annihilator ideal of the radical $\bar{N}$ in $\bar{A}$, and denote by P the two-sided ideal in A consisting of those elements whose residue classes (mod $\mathfrak{z}$) lie in $\bar{P}$. We consider further $r(\mathfrak{z})$ and $r(P)$ in A. The latter is the intersection of all maximal right subideals of the former, because P is the sum of all those left ideals in which $\mathfrak{z}$ is a maximal left subideal.

There exists, by definition, a permutation $\{\nu(\rho)\}$ of $\{\rho\}$ such that $\bar{e}_\rho\bar{P} \cong \bar{e}_{\nu(\rho)}\bar{A}$ ($\cong \bar{e}_{\nu(\rho)}\bar{A}/\bar{e}_{\nu(\rho)}\bar{N}$), $\bar{P}\bar{e}_{\nu(\rho)} \cong \bar{A}\bar{e}_\rho$ and $f(\nu(\rho)) = f(\rho)$. We have $\bar{E}_\rho\bar{P} = \bar{P}\bar{E}_{\nu(\rho)}$, or, $E_\rho P \cup \mathfrak{z} = PE_{\nu(\rho)} \cup \mathfrak{z}$. Now, $\mathfrak{z}$ coincides with the intersection $\bigcap_\rho(\sum_{\kappa\neq\rho} E_\kappa P \cup \mathfrak{z})$, ρ running over $\{\rho\}$. Hence

$$r(\mathfrak{z}) = \sum_\rho r(\sum_{\kappa\neq\rho} E_\kappa P \cup \mathfrak{z}); \tag{13}$$

the summands are two-sided ideals since $\sum_{\kappa\neq\rho} E_\kappa P \cup \mathfrak{z}$ are such. Furthermore, $P/\sum_{\kappa\neq\rho} E_\kappa P \cup \mathfrak{z} \cong \bar{E}_\rho\bar{P}$ is, for each ρ, a direct sum of $f(\nu(\rho)) = f(\rho)$ simple A-left-submoduli isomorphic to $\bar{A}\bar{e}_\rho$, and therefore $r(\sum_{\kappa\neq\rho} E_\kappa P \cup \mathfrak{z})/r(P)$ is a

[36] Observe that bA is two-sided.

[37] This follows also from Theorem 8.

direct sum of $f(\rho)$ simple A-right-submoduli isomorphic to $\bar{e}_{\pi(\rho)}\bar{A}$ (Cf. Theorem 9). But

$$(14)\quad \begin{aligned} r(\textstyle\sum_{\kappa\neq\rho}E_\kappa P \cup \mathfrak{z}) &= r(\textstyle\sum_{\kappa\neq\nu(\rho)}PE_\kappa \cup \mathfrak{z}) = r((P \cap A\textstyle\sum_{\kappa\neq\nu(\rho)}E_\kappa) \cup \mathfrak{z}) \\ &= (r(P) \cup E_{\nu(\rho)}A) \cap r(\mathfrak{z}) = r(P) \cup (E_{\nu(\rho)}A \cap r(\mathfrak{z}))^{38} = E_{\nu(\rho)}r(\mathfrak{z}) \cup r(P), \end{aligned}$$

whence

$$r(\textstyle\sum_{\kappa\neq\nu(\rho)}E_\kappa P \cap \mathfrak{z})/r(P) = E_{\nu(\rho)}r(\mathfrak{z}) \cup r(P)/r(P) = \sum_{i=1}^{f(\nu(\rho))}(e_{\nu(\rho),i}r(\mathfrak{z}) \cup r(P)/r(P)).$$

Hence we find that the right moduli $e_{\nu(\rho),i}r(\mathfrak{z}) \cup r(P)/r(P)$ are simple and $\cong \bar{e}_{\pi(\rho)}\bar{A}$.

According to this fact, we now take for each ρ, $i(i = 1, 2, \cdots, f(\rho))$ an element $b_{\rho,i}$ in $e_{\nu(\rho),i}r(\mathfrak{z})e_{\pi(\rho),i}$ which is not contained in $r(P)$; observe that $f(\pi(\rho)) = f(\rho) = f(\nu(\rho))$. Then $b_{\rho,i}A \cup r(P) = e_{\nu(\rho),i}r(\mathfrak{z}) \cup r(P)$. Moreover, if we put $b = \sum_{\rho,i}b_{\rho,i}$, then $be_{\pi(\rho),i} = b_{\rho,i}$, and therefore

$$bA \cup r(P) = \textstyle\sum_{\rho,i}(e_{\nu(\rho),i}r(\mathfrak{z})) \cup r(P) = \sum_\rho(E_{\nu(\rho)}r(\mathfrak{z})) \cup r(P) = r(\mathfrak{z})$$

(cf. (13), (14)). However, $r(P)$ is the intersection of all maximal right subideals of $r(\mathfrak{z})$, as we observed before. Hence necessarily $bA = r(\mathfrak{z})$.

Similarly $l(\mathfrak{z})$ is a principal left ideal, $l(\mathfrak{z}) = Ac$, and this completes the proof.

THEOREM 16. *Let a ring A satisfy (not only the minimum condition but) also the maximum condition for left and right ideals. In order that every residue class ring of A be Frobeniusean, it is necessary and sufficient that every two-sided ideal $\mathfrak{z}$ in A can be expressed as $\mathfrak{z} = Ac = cA$ $(c \in A)$.*

PROOF. a) Assume that every residue class ring of A is Frobeniusean. Then in particular A is Frobeniusean, and the above theorem 15 tells that every two-sided ideal $\mathfrak{z}(= r(l(\mathfrak{z}))$ is a principal right ideal $\mathfrak{z} = cA$. Moreover $d_r(\mathfrak{z}) = d_r(A) - d_r(A/\mathfrak{z}) = d_l(A) - d_l(A/\mathfrak{z}) = d_l(\mathfrak{z})$, for both A and $A/\mathfrak{z}$ are Frobeniusean.[39] But $d_r(\mathfrak{z}) = d_r(cA) = d_l(Ac)$ by Theorem 8. Hence $d_l(\mathfrak{z}) = d_l(Ac)$. Since $Ac \subseteq \mathfrak{z}$ we have $Ac = \mathfrak{z}$.

b) Assume next that the condition of the theorem is satisfied. The existence of a unit element in A can be seen in the same way as in I, Th. 10.[40] Let $\mathfrak{z} = Ac = cA$. The left ideal $\mathfrak{z} = Ac$ is isomorphic to $A/l(c) = A/l(cA) = A/l(\mathfrak{z})$, whence $d_l(\mathfrak{z}) = d_l(A/l(\mathfrak{z}))$. Since this is the case for every two-sided ideal, we have $r(l(\mathfrak{z})) = \mathfrak{z}$. For, $l(r(l(\mathfrak{z}))) = l(\mathfrak{z})$ whence $d_l(r(l(\mathfrak{z}))) = d_l(A/l(r(l(\mathfrak{z})))) = d_l(A/l(\mathfrak{z})) = d_l(\mathfrak{z})$, which, together with $r(l(\mathfrak{z})) \supseteq \mathfrak{z}$, implies $r(l(\mathfrak{z})) = \mathfrak{z}$. Similarly $d_r(\mathfrak{z}) = d_r(A/r(\mathfrak{z}))$ and $l(r(\mathfrak{z})) = \mathfrak{z}$. We find therefore, according to Theorem 11, that A is Frobeniusean. Our assertion that every residue class ring of A is Frobeniusean is now an immediate consequence of Theorem 15.

[38] Modular law.

[39] See the footnote 21, or Th. 8.

[40] Here we use the composition lengths, and that is the reason that we assumed the maximum condition too.

9. Appendix: Generalized uni-serial rings

As a generalization of the notion of Köthe's *uni-serial* ring,[41] we introduce the following

DEFINITION. We call A a *generalized uni-serial ring*, if A has a unit element and if every left ideal Ae as well as every right ideal eA generated by a primitive idempotent element e possesses only one composition series.[42]

The connection of this notion to our study is that a ring in which every residue class ring is Frobeniusean, as in the case of Theorem 16, is certainly a generalized uni-serial ring, although the converse is not true.

Now, consider a generalized uni-serial ring A. Let $N^{\rho-1} \neq 0$, $N^\rho = 0$, where N is, as before, the radical of A. Let $\sigma(\kappa)$ be, for each κ, an integer such that $e_\kappa N^{\sigma(\kappa)-1} \neq 0$, $e_\kappa N^{\sigma(\kappa)} = 0$; hence $\rho = \mathrm{Max}_\kappa(\sigma(\kappa))$. Since $e_\kappa A$ has only one composition series, every $e_\kappa N^i/e_\kappa N^{i+1}$ $(i = 0, 1, \cdots, \sigma(\kappa) - 1)$ is simple. We put

$$e_\kappa N^i/e_\kappa N^{i+1} \cong \bar{e}_{\varphi(\kappa,i)}\bar{A}.$$

Thus $e_\kappa N^i e_{\varphi(\kappa,i)} \not\subseteq N^{i+1}$ and if d is an element in $e_\kappa N^i e_{\varphi(\kappa,i)}$ not contained in N^{i+1} then $dA = e_\kappa N^i$. Let now $i + j \leqq \sigma(\kappa) - 1$. Then $e_\kappa N^{i+j} = e_\kappa N^i e_{\varphi(\kappa,i)} A N^j = e_\kappa N^i e_{\varphi(\kappa,i)} N^j$. Hence $j \leqq \sigma(\varphi(\kappa, i)) - 1$. Moreover, since $e_{\varphi(\kappa,i)} N^j = e_{\varphi(\kappa,i)} N^j e_{\varphi(\varphi(\kappa,i),j)} A$, we have $e_\kappa N^{i+j+1} \subset e_\kappa N^{i+j} = e_\kappa N^i e_{\varphi(\kappa,i)} N^j e_{\varphi(\varphi(\kappa,i),j)} A \subseteq e_\kappa N^{i+j} e_{\varphi(\varphi(\kappa,i),j)} A$. Therefore

$$\varphi(\kappa, i + j) = \varphi(\varphi(\kappa, i), j). \tag{15}$$

We put similarly $N^{\tau(\kappa)-1} e_\kappa \neq 0$, $N^{\tau(\kappa)} e_\kappa = 0$ and

$$N^i e_\kappa / N^{i+1} e_\kappa \cong \bar{A}\bar{e}_{\psi(\kappa,i)}$$

for $i \leqq \tau(\kappa) - 1$. Then

$$\psi(\kappa, i + j) = \psi(\psi(\kappa, i), j) \tag{16}$$

provided $i + j \leqq \tau(\kappa) - 1$. Further, if $i \leqq \tau(\kappa) - 1$ then $i \leqq \sigma(\psi(\kappa, i)) - 1$ and $\varphi(\psi(\kappa, i), i) = \kappa$, because $e_{\psi(\kappa,i)} N^i e_\kappa \not\subseteq N^{i+1}$. Hence, if $j \leqq \tau(\kappa) - 1$, $i \leqq \sigma(\psi(\kappa, j)) - 1$, then

$$\varphi(\psi(\kappa,j),i) = \begin{cases} \varphi(\varphi(\psi(\kappa, j), j), i - j) = \varphi(\kappa, i - j) \text{ or} \\ \varphi(\psi(\psi(\kappa, j - i), i), i) = \psi(\kappa, j - i) \end{cases} \tag{17}$$

according as $i \geqq j$ or $i \leqq j$.[43]

After this preparation we now proceed in a similar manner as in Köthe (9). Suppose that a left module $\mathfrak{M}$ of our A is a (restricted) direct sum $\sum_s \mathfrak{M}_s$ of (an arbitrary number of) submoduli $\mathfrak{M}_s$ and that each $\mathfrak{M}_s$ is homomorphic to

[41] Köthe (9).

[42] See the footnotes 40 and 41 of Part I.

[43] We obtain a similar formula by interchanging σ, φ with τ, ψ. In particular, $\varphi(\psi(\kappa, j), i) = \psi(\varphi(\kappa, i), j)$ provided $j \leqq \tau(\kappa) - 1$, $i \leqq \sigma(\psi(\kappa, j)) - 1$, $i \leqq \sigma(\kappa) - 1$, $j \leqq \tau(\varphi(\kappa, i)) - 1$.

one of Ae_κ. Let $\mathfrak{M}_s$ be homomorphic to $Ae_{\kappa(s)}$ and let $e_{\kappa(s)}$ correspond to u_s in $\mathfrak{M}_s$; we have $\mathfrak{M}_s = Au_s = Ae_{\kappa(s)}u_s$. Denote further the composition length of $\mathfrak{M}_s$ by $i(s)$. Then $\mathfrak{M}_s \cong Ae_{\kappa(s)}/N^{i(s)}e_{\kappa(s)}$, and $N^{i(s)-1}u_s \neq 0$ while $N^{i(s)}u_s = 0$. Let m be the smallest of all $i(s)$.

LEMMA 4. *If u ($\neq 0$) is an element in $e_\kappa\mathfrak{M}$ such that the composition length l of Au is not greater than m, then there exists an element v in $e_{\varphi(\kappa,m-l)}\mathfrak{M}$ such that $u \in e_\kappa N^{m-l}v$ and the composition length of Av is exactly m.*

PROOF. Suppose for instance $u = a_1u_1 + a_2u_2 + \cdots + a_tu_t$, where $a_s \in Ae_{\kappa(s)}$ and $a_s \neq 0$ $(s = 1, 2, \cdots, t)$. Since $u = e_\kappa u = e_\kappa a_1u_1 + \cdots + e_\kappa a_tu_t$, we can, and shall, assume $a_s \in e_\kappa Ae_{\kappa(s)}$. Let $a_s \in e_\kappa N^{l(s)}e_{\kappa(s)}$ but $\notin e_\kappa N^{l(s)+1}e_{\kappa(s)}$. Then

$$(18) \qquad \kappa = \psi(\kappa(s), l(s)), \qquad \kappa(s) = \varphi(\kappa, l(s)),$$

and $Aa_s = N^{l(s)}e_{\kappa(s)}$. On the other hand $N^lu = 0$ and $N^{l-1}u \neq 0$. From $N^ja_su_s = N^jAa_su_s = N^{j+l(s)}e_{\kappa(s)}u_s = N^{j+l(s)}u_s$ $(s = 1, 2, \cdots, t)$, it follows that $l + l(s) \geqq i(s)$ for all $s = 1, 2, \cdots, t$ while $l + l(s) \leqq$ (whence $=$) $i(s)$ for at least one s.

Take an element d such that $d \in e_\kappa N^{m-l}e_{\varphi(\kappa,m-l)}$ but $\notin N^{m-l+1}$. Then[44] $a_s \in e_\kappa N^{l(s)}e_{\kappa(s)} = e_\kappa N^{m-l}N^{l(s)-m+l}e_{\kappa(s)} = dN^{l(s)-m+l}e_{\kappa(s)}$, whence $a_s = db_s$ with $b_s \in e_{\varphi(\kappa,m-l)}N^{l(s)-m+l}e_{\kappa(s)}$. Here $\varphi(\varphi(\kappa, m - l), l(s) - m + l) = \varphi(\kappa, m - l + l(s) - m + l) = \varphi(\kappa, l(s)) = \kappa(s)$ according to (15), (18), and $Ab_s = N^{l(s)-m+l}e_{\kappa(s)}$.

Put now $v = b_1u_1 + b_2u_2 + \cdots + b_tu_t$. Then $u = dv$ and $e_{\varphi(\kappa,m-l)}v = v$. $N^mv = 0$, because $l + l(s) \geqq i(s)$ whence $N^mb_s \subseteq N^mN^{l(s)-m+l}e_{\kappa(s)} = N^{l(s)+l}e_{\kappa(s)} = 0$ $(s = 1, 2, \cdots, t)$. But $N^{m-1}v \neq 0$, since $N^{m-1}b_s = N^{l(s)+l-1}e_{\kappa(s)}$ and $l(s) + l \leqq i(s)$ for at least one s. Hence the composition length of Av is exactly m.

LEMMA 5. *Let $\mathfrak{M}$ be the same as above. Suppose moreover that $\mathfrak{M}$ is contained in an over (A-left-) module $\mathfrak{m}$ and that $\mathfrak{m}$ is generated by $\mathfrak{M}$ and a cyclic module $\mathfrak{M}'$ which is homomorphic to one of Ae_κ and whose composition length does not exceed m. Then $\mathfrak{m}$ is a direct sum $\mathfrak{M} + \mathfrak{M}''$ of $\mathfrak{M}$ and a second module $\mathfrak{M}''$ homorphic to $\mathfrak{M}'$.*

PROOF. Let $\mathfrak{M}' = Au \cong Ae_\kappa/N^he_\kappa$, $u = e_\kappa u$, and put $\mathfrak{U} = \mathfrak{M} \cap \mathfrak{M}'$. If $\mathfrak{U} = 0$, then we are already through. Suppose $\mathfrak{U} \neq 0$ and let $\mathfrak{U} = N^lu = N^le_\kappa u$; the composition length of $\mathfrak{U}$ is then $h - l$. Take an element c in $e_{\psi(\kappa,l)}N^le_\kappa$ not contained in N^{l+1}. Then $cu \in \mathfrak{U}$ and $Acu = N^lu = \mathfrak{U}$. But Acu is homomorphic to $Ae_{\psi(\kappa,l)}$. Hence there exists, by the above lemma 4, an element v in $\mathfrak{M}$ such that $v = e_\lambda v$, $Av \cong Ae_\lambda/N^me_\lambda$ and $cu \in e_{\psi(\kappa,l)}N^{m-(h-l)}v$, where $\lambda = \varphi(\psi(\kappa, l), m - (h - l)) = \varphi(\kappa, m - h)$.[45] But $e_{\psi(\kappa,l)}N^{m-(h-l)}e_\lambda = e_{\psi(\kappa,l)}N^lN^{m-h}e_\lambda = cN^{m-h}e_\lambda = ce_\kappa N^{m-h}e_\lambda$. Therefore $cu = cc'v$ with an element c' *in* $e_\kappa N^{m-h}e_\lambda$.

We put now $u'' = u - c'v$. Obviously $\mathfrak{m} = \mathfrak{M} \cup Au''$. Moreover, $e_\kappa u'' = e_\kappa u - e_\kappa c'v = u - c'v = u''$, whence Au'' is homomorphic to Ae_κ. Since $cu'' = cu - cc'v = 0$, we have $N^lu'' = Acu'' = 0$. Furthermore $\mathfrak{M} \cap Au'' = 0$, that is, $\mathfrak{m} = \mathfrak{M} + Au''$. For, if $au'' = w \in \mathfrak{M}(a \in Ae_\kappa)$, then $au - ac'v = w$, $au = w + ac'v \in \mathfrak{M}$. Hence $au \in \mathfrak{M} \cap Au = \mathfrak{U}$, and therefore $a \in N^le_\kappa$, $au'' = 0$. This completes our proof.

[44] Observe that $l + l(s) \geqq i(s) \geqq m$.

[45] Cf. (17).

Now, the following theorem, which is an extension of the main theorem in Köthe (9), follows readily from the above lemma:

THEOREM 17.[46] *Let A be a generalized uni-serial ring. Then every left module of A is a (restricted) direct sum of cyclic submoduli, each of which is homomorphic to one of Ae_κ. The similar situation prevails for a right module of A.*

REMARK. In the special case where A is, as in Part I, an *algebra*, the converse is also true. Namely, an algebra which possesses the property of Theorem 17 is necessarily a generalized uni-serial algebra. For, if a left module $N^{i-1}e/N^ie$, where e is a primitive idempotent element, is not simple, then a right module which belongs to a representation of A equivalent to the one defined by the left module Ae/N^ie is homomorphic to none of $e_\kappa A$, whereas it is certainly directly indecomposable.[47]

Added in proof (Oct. 22, 1940): The remark adjoining the definition of Frobeniusean rings (Chapter II, 4) ought to have been placed after Theorem 6. In the proof of the theorem we showed $l(N) \subseteqq r(N)$ using only the condition i), and from this follows the relation $e_\kappa r(N) \supseteqq e_\kappa l(N) = \mathfrak{r}_\kappa$ used in the remark.

Asano (2), Theorem 2 and our Theorem 16 express essentially one and the same fact (Cf. Part I, Theorem 10 too). Namely, a ring whose residue class rings are all Frobeniusean is uni-serial. Further, the converse of Theorem 17 is valid generally. For all this cf. the writer's *Note on uni-serial and generalized uni-serial rings*, Proc. Imp. Acad. Tokyo Vol. XVI (1940) p. 285.

INSTITUTE FOR ADVANCED STUDY

BIBLIOGRAPHY (continued from Part I)

(18) R. BRAUER, *Über die Kleinsche Theorie der algebraischen Gleichungen*, Math. Annalen 110 (1934).

(19) E. CARTAN, *Leçons sur les invariants intégraux.*

(20) M. DEURING, *Galoissche Theorie und Darstellungstheorie*, Math. Annalen 107 (1932).

(21) M. DEURING, *Anwendungen der Darstellungen von Gruppen durch lineare Substitutionen auf die Galoissche Theorie*, Math. Annalen 113 (1936).

(22) J. HAANTJES, *Halblineare Transformationen*, Math. Annalen 114 (1937).

(23) C. HOPKINS, *Nil-rings with minimal condition for admissible left ideals*, Duke Math. J. 4 (1938).

(24) C. HOPKINS, *Rings with minimal condition for left ideals*, Ann. of Math. 40 (1939).

(25) T. NAKAYAMA, *Über die Klassifikation der halblinearen Transformationen*, Proc. Phys.-Math. Soc. Japan 19 (1937).

(26) E. NOETHER, *Normalbasis bei Körpern ohne höhere Verzweigung*, Crelles Journal 167 (1931).

(27) I. SCHUR, *Einige Bemerkungen zur vorstehenden Arbeit des Herrn Speiser*, Math. Zeitschr. 5 (1919).

(28) R. STAUFFER, *The construction of a normal basis in a separable normal extension field*, American J. Math. 58 (1936).

(29) B. L. VAN DER WAERDEN, *Gruppen von linearen Transformationen*, Ergebn. Math. 4 (1935).

(30) H. WEYL, *Generalized Riemann matrices and factor sets*, Ann. Math. 37 (1936).

[46] Cf. I, Th. 11.

[47] Cf. the proof of I, Th. 11 and I, footnote 40.

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

ÜBER DIE TOPOLOGIE DER GRUPPEN-MANNIGFALTIGKEITEN UND IHRE VERALLGEMEINERUNGEN

BY HEINZ HOPF

Received* by Compositio Math., August 23, 1939

Einleitung.[1]

1. In der geschlossenen und orientierbaren Mannigfaltigkeit M sei eine "stetige Multiplikation" erklärt, das heißt: jedem geordneten Punktepaar (p, q) von M ist als "Produkt" ein Punkt pq von M zugeordnet, der stetig von dem Paar (p, q) abhängt. Setzen wir

$$pq = l_p(q),$$

so ist l_p bei festem p und variablem q eine Abbildung von M in sich; die Abbildungen l_p hängen stetig von dem Parameter p ab, und sie haben daher alle den gleichen Abbildungsgrad c_l. Analog ist der Grad c_r der Abbildungen r_q bestimmt, die durch

$$pq = r_q(p)$$

gegeben sind.

Definiert man etwa die stetige Multiplikation so, daß pq für alle (p, q) ein fester Punkt von M ist, so ist $c_l = c_r = 0$; setzt man $pq = p$ oder setzt man $pq = q$ für alle (p, q), so ist $c_l = 0$, $c_r = 1$ bzw. $c_l = 1$, $c_r = 0$. Diese trivialen stetigen Multiplikationen sind in jeder Mannigfaltigkeit möglich; dagegen kann man, wie sich zeigen wird, nur in sehr speziellen Mannigfaltigkeiten stetige Multiplikationen so definieren, daß

$$c_l \neq 0 \qquad \textit{und} \qquad c_r \neq 0$$

ist. Eine Mannigfaltigkeit,[2] welche eine solche Multiplikation zuläßt, soll eine Γ-*Mannigfaltigkeit* heißen.[2a]

Der Begriff der Γ-Mannigfaltigkeit ist eine Verallgemeinerung des Begriffes der *Gruppen*-Mannigfaltigkeit; ist nämlich M eine Gruppen-Mannigfaltigkeit, d.h. ist in M eine stetige Multiplikation erklärt, welche die Gruppen-Axiome erfüllt, so ist für den Punkt e, welcher die Gruppen-Eins darstellt, sowohl die

---

* *Editors' Note.* This paper was originally submitted to *Compositio Mathematica*, August 23, 1939. It was transferred to the *Annals of Mathematics* (received November 18, 1940) after the *Compositio Mathematica* ceased publication.

[1] Eine kurze Ankündigung dieser Arbeit ohne Beweise ist in den C. R. 208 (1939), 1266–1267, erschienen.

[2] Unter einer "Mannigfaltigkeit" ist in dieser Arbeit immer eine *geschlossene und orientierbare* Mannigfaltigkeit zu verstehen.

[2a] Die Gültigkeit des assoziativen Gesetzes wird also nicht gefordert.

Abbildung l_e als auch die Abbildung r_e die Identität von M, und daher ist $c_l = c_r = 1$.

Somit gelten alle Sätze, die für Γ-Mannigfaltigkeiten bewiesen werden, insbesondere für geschlossenen Gruppen-Mannigfaltigkeiten.[3]

2. Wir werden Homologie-Eigenschaften von Mannigfaltigkeiten untersuchen; dabei soll *als Koeffizientenbereich der Körper der rationalen Zahlen* dienen.[4] Wie üblich fassen wir die Homologieklassen einer Mannigfaltigkeit M zu dem Homologie-Ring $\mathfrak{R}(M)$ zusammen: in ihm ist die Addition die der Bettischen Gruppen, und die Multiplikation ist durch die Schnitt-Bildung erklärt. Infolge der Benutzung rationaler Koeffizienten entgehen uns zwar gewisse Feinheiten der Struktur von M, so die etwa vorhandene Torsion; immerhin stimmen zwei Mannigfaltigkeiten M_1, M_2, deren rationalen Homologie-Ringe $\mathfrak{R}(M_1)$ und $\mathfrak{R}(M_2)$ einander dimensionstreu isomorph sind, in den wichtigsten algebraisch-topologischen Eigenschaften überein, insbesondere in den Werten der Bettischen Zahlen.

Unser Hauptziel ist der Beweis des folgenden Satzes:

Satz I. *Der Homologie-Ring $\mathfrak{R}(\Gamma)$ einer Γ-Mannigfaltigkeit Γ ist dimensionstreu isomorph dem Homologie-Ring $\mathfrak{R}(\Pi)$ eines topologischen Produktes*

$$\Pi = S_{m_1} \times S_{m_2} \times \cdots \times S_{m_l}, \qquad l \geqq 1,$$

in welchem S_m die m-dimensionale Sphäre bezeichnet und alle Dimensionszahlen $m_1, m_2, \cdots, m_l$ ungerade sind.

3. Da man die Struktur der Ringe $\mathfrak{R}(\Pi)$ vollständig übersieht, kann man den Inhalt des Satzes I auch durch eine ausfürhliche Beschreibung der Struktur der Ringe $\mathfrak{R}(\Gamma)$ ausdrücken. Hierfür machen wir noch die folgenden terminologischen Bemerkungen:

Der Ring $\mathfrak{R}(M)$ einer beliebigen n-dimensionalen Mannigfaltigkeit[2] M enthält ein Eins-Element: es wird durch den orientierten n-dimensionalen Grundzyklus von M dargestellt; wir bezeichnen es durch 1. Die Dimension eines Elementes z von $\mathfrak{R}(M)$ nennen wir $d(z)$; daneben betrachten wir häufig die "duale Dimension" $\delta(z) = n - d(z)$. Unter einer "vollen additiven Basis" von $\mathfrak{R}(M)$ verstehen wir die Vereinigung von Homologie-Basen der Dimensionen $0, 1, \cdots, n$.

Nun läßt sich der Satz I folgendermaßen aussprechen:

[3] Die Topologie der Gruppen-Mannigfaltigkeiten wird in den folgenden beiden Schriften von E. Cartan behandelt: (a) *La Théorie des Groupes Finis et Continus et l'Analysis Situs* [Paris 1930, Mémorial Sc. Math. XLII]; (b) *La Topologie des Groupes de Lie* [Paris 1936, Actualités Scient. et Industr. **358**; sowie: L'Enseignement math. **35** (1936), 177–200; sowie: Selecta, Jubilé Scientifique, Paris 1939, 235–258].

[4] Tatsächlich werden wir von dem Koeffizientenbereich nur benutzen, daß er ein *Körper der Charakteristik* 0 ist.

Satz I (2. *Fassung*). *Aus dem Ringe* $\mathfrak{R}(\Gamma)$ *einer* Γ-*Mannigfaltigkeit* Γ *lassen sich Elemente* $z_1, z_2, \cdots, z_l$ *so auswählen, daß die* 2^l *Elemente*

$$1 \quad ; \quad z_i \quad ; \quad z_{i_1}\cdot z_{i_2} \quad (i_1 < i_2) \quad ;$$

$$z_{i_1}\cdot z_{i_2}\cdot z_{i_3} \quad (i_1 < i_2 < i_3) \quad ; \quad \cdots \quad ; \quad z_1\cdot z_2\cdot \cdots \cdot z_l$$

eine volle additive Basis bilden; die Multiplikation in $\mathfrak{R}(\Gamma)$ *ist durch das distributive Gesetz, das assoziative Gesetz und die antikommutative Regel*

$$z_j\cdot z_i = -z_i\cdot z_j$$

— *in welcher speziell die Regel*

$$z_i\cdot z_i = 0$$

enthalten ist—vollständig bestimmt; kurz: $\mathfrak{R}(\Gamma)$ *ist der Ring der (inhomogenen) Multilinearformen in den antikommutativen Größen* $z_1, z_2, \cdots, z_l$ *mit rationalen Koeffizienten. Alle* z_i *sind homogendimensional,*[5] *und ihre dualen Dimensionen* $\delta(z_i) = m_i$ *sind ungerade; die Dimension von* Γ *ist*

$$n = m_1 + m_2 + \cdots + m_l,$$

und allgemein ist

$$d(z_{i_1}\cdot z_{i_2}\cdot \cdots \cdot z_{i_r}) = n - m_{i_1} - m_{i_2} - \cdots - m_{i_r}.$$

Daß hierdurch gerade das Bestehen eines dimensionstreuen Isomorphismus zwischen $\mathfrak{R}(\Gamma)$ und $\mathfrak{R}(\Pi)$ ausgedrückt wird, erhellt aus der folgenden Tatsache, die man leicht bestätigt: bezeichnet man den orientierten Grundzyklus von S_m selbst mit S_m und mit p immer einen einfach gezählten Punkt, so besitzen in $R(\Pi)$ die Elemente

$$(*) \qquad \begin{aligned} Z_1 &= p \times S_{m_2} \times S_{m_3} \times \cdots \times S_{m_l}, \\ Z_2 &= S_{m_1} \times p \times S_{m_3} \times \cdots \times S_{m_l}, \\ &\cdots\cdots\cdots\cdots\cdots \\ Z_l &= S_{m_1} \times S_{m_2} \times S_{m_3} \times \cdots \times p \end{aligned}$$

genau die analogen Eigenschaften, welche wir in $\mathfrak{R}(\Gamma)$ soeben den Elementen $z_1, z_2, \cdots, z_l$ zugeschrieben haben.

Aus der zweiten Fassung des Satzes I liest man unter anderem die folgende bemerkenswerte Eigenschaft der Γ-Mannigfaltigkeiten ab:

Satz Ib. *Jedes homogen-dimensionale Element* z *von* $\mathfrak{R}(\Gamma)$, *für welches die*

[5] Ein Komplex K heißt homogen-dimensional von der Dimension r, wenn jedes Simplex von K auf einem r-dimensionalen Simplex von K liegt. Ein Element des Homologie-Ringes $\mathfrak{R}(M)$ heißt homogen-dimensional, wenn es in M durch einen homogen-dimensionalen Zyklus repräsentiert wird. Die additive Gruppe von $\mathfrak{R}(M)$ ist die direkte Summe der Gruppen der homogen r-dimensionalen Elemente mit $r = 0, 1, \cdots, n$.

duale Dimension $\delta(z)$ gerade und positiv ist, läßt sich durch Multiplikation und Addition aus höherdimensionalen Elementen erzeugen.

4. Der Satz I enthält weitgehende Aussagen über die Bettischen Zahlen einer Γ-Mannigfaltigkeit. Unter dem "Poincaréschen Polynom" eines Komplexes K verstehen wir das Polynom

$$P_K(t) = p_0 + p_1 t + p_2 t^2 + \cdots$$

in einer Unbestimmten t, wobei der Koeffizient p_r die r-te Bettische Zahl von K ist. Für die Sphäre S_m ist

$$P_{S_m}(t) = 1 + t^m;$$

bei der Bildung des topologischen Produktes $K_1 \times K_2$ zweier Komplexe K_1 und K_2 gilt nach der Formel von Künneth[6] die Regel

$$P_{K_1 \times K_2}(t) = P_{K_1}(t) \cdot P_{K_2}(t);$$

daher ist in dem Satz I (Nr. 2) der folgende Satz enthalten:

SATZ I'. *Das Poincarésche Polynom einer Γ-Mannigfaltigkeit Γ hat die Gestalt*

$$P_\Gamma(t) = (1 + t^{m_1}) \cdot (1 + t^{m_2}) \cdot \cdots \cdot (1 + t^{m_l}), \tag{1}$$

wobei alle Exponenten m_i ungerade sind.

Wir heben einige der zahlreichen Beziehungen zwischen den Bettischen Zahlen hervor, die sich aus (1) ablesen lassen:

(a) *Die Eulersche Charakteristik ist* 0;[7]

denn die Charakteristik eines Komplexes K ist gleich $P_K(-1)$.

(b) *Die Summe der Bettischen Zahlen ist eine Potenz von* 2;[8]

denn diese Summe ist für einen Komplex K gleich $P_K(+1)$.

Γ sei n-dimensional; dann ist $p_n = 1$, $p_r = 0$ für $r > n$, also n der Grad von $P_\Gamma(t)$ und

$$m_1 + m_2 + \cdots + m_l = n.$$

Da der r-te Koeffizient des Polynoms (1) offenbar nicht größer ist als der r-te Koeffizient des Polynoms

$$(1 + t)^n = (1 + t)^{m_1} \cdot (1 + t)^{m_2} \cdots (1 + t)^{m_l},$$

so sieht man:

(c) *Es ist* $p_r \leqq \binom{n}{r}$ *für alle* r.[9]

---

[6] Alexandroff-Hopf, Topologie I (Berlin 1935), 309, Formel (13').

[7] Falls die Abbildungen l_p und r_q *topologisch* sind—also insbesondere, falls Γ eine *Gruppen*-Mannigfaltigkeit ist—, wird für $p_1 \neq p_2$ durch $f(q) = l_{p_1}^{-1} l_{p_2}(q)$ eine Abbildung von Γ auf sich erklärt, welche sich stetig in die Identität deformieren läßt und keinen Fixpunkt besitzt; dann folgt der obige Satz aus einem bekannten Fixpunktsatz.

[8] Für Gruppen-Mannigfaltigkeiten: Cartan[3] (b), 24.

[9] Für Gruppen-Mannigfaltigkeiten: H. Weyl, The classical groups [Princeton 1939], 279, als Korollar eines Satzes von Cartan (cf.[13]).

Man kann (1) in der Form

$$(1') \qquad P_\Gamma(t) = (1+t)^{l_1}\cdot(1+t^3)^{l_3}\cdot(1+t^5)^{l_5}\cdots, \qquad l_s \geqq 0,$$

schreiben; Ausrechnung ergibt

$$(1'') \quad P_\Gamma(t) = 1 + l_1 t + \binom{l_1}{2}t^2 + \left(l_3 + \binom{l_1}{3}\right)t_3 + \left(l_1 l_3 + \binom{l_1}{4}\right)t^4 + \cdots$$

Es ist also

$$(2) \qquad p_1 = l_1 .$$

Da nun der Koeffizient von t^r in dem Produkt (1') offenbar nicht kleiner ist als in dem Faktor

$$(1+t)^{l_1} = (1+t)^{p_1},$$

so gilt:

(d) *Es ist* $p_r \geqq \binom{p_1}{r}$ *für alle* r.[10]

Nach (1') ist

$$l_1 + 3l_3 + 5l_5 + \cdots = n,$$

also nach (2):

$$p_1 = n - 3l_3 - 5l_5 - \cdots ;$$

daher läßt sich (c) für $r = 1$ verschärfen:

(e) *Es ist entweder* $p_1 = n$ *oder* $p_1 = n - 3$ *oder* $p_1 \leqq n - 5$.[11]

Ferner liest man aus (1'') und (2) die folgende Verschärfung von (d) für $r = 2$ ab:

(f) *Es ist* $p_2 = \binom{p_1}{2}$,

also speziell:

(f_0) *Ist* $p_1 = 0$ *oder* $p_1 = 1$, *so ist* $p_2 = 0$.[12]

Ebenso sieht man aus (1'') und (2):

(g) *Ist* $p_1 = 0$, *so ist auch* $p_4 = 0$.

Man kann ohne Mühe noch eine Reihe ähnlicher Relationen feststellen, z.B. die folgenden:

---

[10] Für Gruppen-Mannigfaltigkeiten wie[9]; für wesentlich allgemeinere Räume mit stetiger Multiplikation: W. Hurewicz [Proc. Akad. Amsterdam **39** (1936), 215-224].

[11] Die Relation $p_1 \leqq n$ wurde für verallgemeinerte Gruppenräume zuerst von P. Smith [Annals of Math. (2) **36** (1935), 210-229] und dann von Hurewicz als Korollar aus dem unter[10] zitierten Satz bewiesen.

[12] Wegen der zweiten Bettischen Zahl einer Gruppen-Mannigfaltigkeit vergl. man Cartan[3], (b), 14 und 23-24.

(h) *Es sei* $p_1 = 0$; *dann ist*

$$3p_3 + 5p_5 + 7p_7 \leqq n,$$

$$p_{ir} \geqq \binom{p_i}{r} \text{ für } i = 3, 5, 7 \text{ und beliebiges } r.$$

5. Für *Liesche Gruppen* kann der Satz I, auf Grund von Sätzen, die wir E. Cartan und G. de Rham verdanken, aus der Sprache der Homologie-Theorie in die Sprache der Theorie ddr invarianten Differentialformen übersetzt werden.[13] Ich begnüge mich mit der Formulierung des Ergebnisses, im Anschluß an die 2. Fassung des Satzes I (Nr. 3):[14]

Die Mannigfaltigkeit G repräsentiere eine geschlossene Liesche Gruppe. Dann kann man aus der Gesamtheit der Differentialformen, welche in G invariant gegenüber den Operationen der Gruppe sind, Formen

$$\omega_1, \omega_2, \cdots, \omega_l,$$

deren Grade

$$m_1, m_2, \cdots, m_l$$

seien, so auswählen, daß sie die folgenden Eigenschaften besitzen:

1) *Für jedes r bilden diejenigen äußeren Produkte*

$$\omega_{i_1} \cdot \omega_{i_2} \cdot \cdots \cdot \omega_{i_\rho},$$

für welche

$$m_{i_1} + m_{i_2} + \cdots + m_{i_\rho} = r, \qquad i_1 < i_2 < \cdots < i_\rho$$

ist, eine lineare Basis (in Bezug auf konstante Koeffizienten) der invarianten Differentialformen des Grades r;

2) *alle* m_i *sind ungerade;*

3) *es ist* $m_1 + m_2 + \cdots + m_l$ *gleich der Dimension von G.*

Hierin ist unter anderem die folgende Tatsache enthalten, die dem Satz Ib (Nr. 3) entspricht:

Jede invariante homogene Differentialform geraden Grades läßt sich aus invarianten Differentialformen kleinerer Grade durch äußere Multiplikation und Addition erzeugen.

6. Auch bei Beschränkung auf Liesche Gruppen sind, soweit ich sehe, sowohl der Satz I als auch der schwächere Satz I′ neu. Allerdings waren diese Sätze bereits für eine so große und wichtige Reihe von Spezialfällen bekannt, daß ihre Gültigkeit für beliebige Liesche Gruppen vermutet werden konnte. L.

---

[13] E. Cartan, Sur les invariants intégraux . . . [Annales Soc. polonaise de Math. **8** (1929), 181–225 (= Selecta, 203–233)]; G. de Rham, Sur l'analysis situs . . . [Journ. de Math. **10** (1931), 115–200].—Man vergl. auch H. Weyl, a.a.O.[9], 276 ff.

[14] Hier muß als Koeffizientenbereich der Körper der *reellen* Zahlen dienen; man vergl.[4]

Pontrjagin, R. Brauer und C. Ehresmann haben nämlich, mit verschiedenen Methoden, die Bettischen Zahlen derjenigen einfachen geschlossenen Lieschen Gruppen bestimmt, welche den vier großen Klassen in der Aufzählung von Killing-Cartan angehören, und diese Methoden liefern nicht nur den Satz I′, sondern auch den Satz I für die genannten Gruppen.[15]

Ausgehend von diesem Resultat könnte man wohl folgendermaßen zu einem Beweis des Satzes I für *alle* geschlossenen Lieschen Gruppen gelangen: Man verifiziere die Gültigkeit des Satzes auch an den fünf einfachen geschlossenen "Ausnahme"-Gruppen in der Killing-Cartanschen Aufzählung; dann übertrage man den—nunmehr für alle *einfachen* geschlossenen Lieschen Gruppen bewiesenen—Satz auf *alle* geschlossenen Lieschen Gruppen, indem man die, aus der Cartanschen Theorie bekannte, Rolle ausnützt, welche die einfachen Gruppen als Bausteine beliebiger Gruppen spielen.

Aber ganz abgesehen von der Frage, ob die direkte Bestätigung des Satzes an den fünf Ausnahme-Gruppen wirklich gelingt, würde ein solcher Beweis aus zwei Gründen nicht vollständig befriedigen. Erstens würde er so umfangreiche und tiefgehende Teile der Theorie der kontinuierlichen Gruppen als Hilfsmittel verwenden, daß dieser Aufwand in keinem rechten Verhältnis zu dem elementar-topologischen Charakter des Satzes selbst stünde. Zweitens würde ein solcher Beweis in einer *Verifizierung* gipfeln; somit würde er zwar besonders konkrete Aufschlüsse über diejenigen *speziellen* Mannigfaltigkeiten liefern, an denen die Verifizierung stattfindet—also über die Mannigfaltigkeiten der einfachen geschlossenen Lieschen Gruppen—, es würde aber wohl doch der Wunsch nach einem Beweis offen bleiben, welcher *allgemeine* Gründe für die Gültigkeit des Satzes erkennen ließe.[16]

Daher glaube ich, daß selbst dann, wenn die direkte Verifizierung des Satzes I an den fünf Ausnahme-Gruppen und damit ein anderer Beweis für alle geschlossenen Lieschen Gruppen gelingt, doch unser Beweis, welcher für alle Γ-Mannigfaltigkeiten gilt und infolgedessen aus der Lieschen Theorie *nichts* benutzt, auch für die Lieschen Gruppen-Mannigfaltigkeiten willkommen ist.

7. Andererseits weiß man, daß der Satz I gewisse Verschärfungen erlaubt, wenn man sich auf Gruppen-Mannigfaltigkeiten beschränkt; dann unterliegen nämlich die Zahlen m_i, die im Satz I auftreten, gewissen Gesetzen; so folgt aus Sätzen von E. Cartan:[17] entweder sind alle $m_i = 1$ — dann ist die Gruppe Abelsch—, oder wenigstens ein m_i ist gleich 3. Diesen Satz oder ähnliche Sätze

---

[15] L. Pontrjagin [C. R. Acad. Sc. U.R.S.S. **1** (1935), 433–437 und C. R. Paris, **200** (1935), 1277–1280].—R. Brauer [C. R. **201** (1935), 419–421] (man vergl. auch H. Weyl, a.a.O.[9], 232 ff.).—C. Ehresmann [C. R. **208** (1939), 321–323; 1263–1265].

[16] Cartan[3], (b), 26: "... Mais même en nous bornant à la simple détermination des nombres de Betti des groupes simples, on ne devra pas s'estimer complètement satisfait si on arrive à faire cette détermination pour les cinq groupes exceptionnels. ... Il faut espérer qu'on trouvera aussi une raison de portée générale expliquant la forme si particulière des polynomes de Poincaré des groupes simples clos."

[17] A.a.O.[3]: (a), 42–43; (b), 24.

mit unserer Methode zu beweisen, welche immer alle Γ-Mannigfaltigkeiten gleichzeitig behandelt, ist prinzipiell unmöglich; denn für Γ-Mannigfaltigkeiten unterliegen die m_i überhaupt keiner Einschränkung; es gilt nämlich der folgende Satz:

SATZ II. *Jedes Sphären-Produkt*

$$S_{m_1} \times S_{m_2} \times \cdots \times S_{m_l}, \qquad l \geqq 1,$$

in welchem die Dimensionszahlen $m_1, m_2, \cdots, m_l$ *ungerade sind, ist eine* Γ-*Mannigfaltigkeit.*

Aus diesem Satz geht hervor, daß der Begriff der Γ-Mannigfaltigkeit nicht nur seiner Definition nach, sondern auch tatsächlich viel allgemeiner ist als der Begriff der Gruppen-Mannigfaltigkeit: nach Satz II sind alle Sphären S_{2k+1} Γ-Mannigfaltigkeiten, während nach einem bekannten, soeben erwähnten Satz von Cartan unter allen Sphären S_n allein S_1 und S_3 Gruppenräume sind.

8. Die Aufgabe, diejenigen Ringe aufzuzählen, welche als Homologie-Ringe von Γ-Mannigfaltigkeiten auftreten, ist durch die Sätze I und II vollständig gelöst.

Die Gültigkeit des Satzes II wird rasch im §1 durch direkte Angabe geeigneter stetiger Multiplikationen bestätigt.

Im §2 werden Erzeugenden-Systeme beliebiger Homologie-Ringe betrachtet. Im Rahmen dieser Betrachtung wird der Satz I in zwei Teile zerlegt—Satz Ia und Satz Ib, von denen wir den zweiten schon in Nr. 3 ausgesprochen haben. Im Satz Ib (Nr. 15) tritt der Begriff des "maximalen" Elementes eines Homologie-Ringes auf, der auch für andere Zwecke als unseren gegenwärtigen wichtig und brauchbar sein dürfte; wir werden sogleich noch auf ihn zurückkommen (Nr. 9).

Der Ansatz zum Beweis der Sätze Ia und Ib, und damit des Satzes I, ist der folgende: man fasse die Punktepaare (p, q) von M als die Punkte $p \times q$ der Produkt-Mannigfaltigkeit $M \times M$ auf; durch eine stetige Multiplikation pq in M, wie wir sie in Nr. 1 erklärt haben, ist dann eine stetige Abbildung F von $M \times M$ in M bestimmt: $F(p \times q) = pq$; diese Abbildungen F sind mit Hilfe des "Umkehrungs-Homomorphismus" zu untersuchen. Entsprechend diesem Ansatz werden zunächst im §3 einige einfache Eigenschaften des Ringes $\mathfrak{R}(M \times M)$ zusammengestellt; sodann wird im §4, nachdem an seinem Anfang kurz an die Theorie des Umkehrungs-Homomorphismus erinnert worden ist, der Beweis der Sätze Ia und Ib geführt.

9. Der schon erwähnte Begriff des maximalen Elementes ist der folgende: ein homogen-dimensionales Element eines Homologie-Ringes $\mathfrak{R}(M)$ heißt maximal, wenn es nicht durch Multiplikation und Addition aus höherdimensionalen Elementen erzeugt werden kann. Im §5 werden die maximalen Elemente noch etwas näher betrachtet, und es werden ihnen jetzt die "minimalen" Elemente gegenübergestellt: das sind diejenigen homogen-dimensionalen Ele-

mente v von $\mathfrak{R}(M)$, welche keine Vielfachen $w = u \cdot v$ mit $0 < d(w) < d(v)$ besitzen. Die Untersuchung führt erstens leicht zu einem Satz über eine gewisse Dualität zwischen den maximalen und den minimalen Elementen (Nr. 33) und zweitens, unter Benutzung des Umkehrungs-Homomorphismus, zu einer bemerkenswerten Invarianz-Eigenschaft der minimalen Elemente (Nr. 34). Diese Tatsachen, zusammen mit dem Satz Ib, liefern noch als Korollar den folgenden Satz, der eine kräftige Verallgemeinerung der Tatsache darstellt, daß eine Sphäre gerader Dimension nicht als Gruppen-Mannigfaltigkeit auftreten kann:

Satz III. *In den Γ-Mannigfaltigkeiten sind die stetigen Bilder von Sphären gerader Dimension immer homolog* 0.

Zum Schluß (Nr. 37) wird ein Problem formuliert, das durch die erwähnte Methode von Pontrjagin[15] angeregt ist und das für die weitere topologische Untersuchung der Gruppen-Mannigfaltigkeiten wichtig sein dürfte; es wird eine Vermutung ausgesprochen, in welcher die minimalen Elemente einer Gruppen-Mannigfaltigkeit eine Hauptrolle spielen.

§1. Beweis des Satzes II.

Der Satz II (Nr. 7) läßt sich in die folgenden beiden Teile zerlegen:

Satz IIa. *Für ungerades m ist die Sphäre S_m eine Γ-Mannigfaltigkeit.*[18]

Satz IIb. *Das topologische Produkt $\Gamma \times \Gamma'$ zweier Γ-Mannigfaltigkeiten Γ und Γ' ist selbst eine Γ-Mannigfaltigkeit.*

10. *Beweis des Saltzes* IIa.[19] Für jeden Punkt q der Sphäre S_m bezeichne r_q die Spiegelung der S_m an demjenigen Durchmesser, auf welchem q liegt; wir setzen $pq = l_p(q) = r_q(p)$.

Die Abbildung r_q ist topologisch, also ist $c_r = \pm 1$ (und zwar, wie man leicht sieht, $c_r = -1$). Wir behaupten weiter: $c_l = \pm 2$ (und zwar ist $c_l = +2$).

p sei ein fester Punkt auf S_m; durch l_p wird jeder Großkreis, auf dem p liegt, auf sich abgebildet, und zwar folgendermaßen: führt man auf dem Kreis eine Winkelkoordinate mit p als Nullpunkt ein, und ist dann q der Punkt mit der Koordinate α, so hat $pq = l_p(q)$ die Koordinate 2α. Daraus ergibt sich: sowohl die offene Halbkugel H von S_m, deren Mittelpunkt p ist, als auch ihre antipodische Halbkugel H' wird topologisch auf $S_m - p'$ abgebildet, wobei p' der Antipode von p ist; die gemeinsame Randsphäre von H und H' geht in den Punkt p' über. Bezeichnen wir mit q' immer den Antipoden von q, so ist der Zusammenhang swischen den Abbildungen der beiden Halbkugeln H und H' durch die Beziehung $l_p(q') = l_p(q)$ gegeben. Nun hat bei ungeradem m die Involution der S_m, welche je zwei Antipoden vertauscht, den Grad $+1$; daher haben, wenn wir die Orientierungen von H und H' durch eine feste Orientierung der S_m festlegen, die topologischen Abbildungen l_p von H und H' den *gleichen*

[18] Daß für *gerades* m die S_m *nicht* Γ-Mannigfaltigkeit ist, ist in Nr. 4 (a) und in Satz III (Nr. 9) enthalten.

[19] Wiedergabe des Beweises von "Satz IV" aus meiner Arbeit in den Fund. Math. **25** (1935), 427–440.

Grad $\epsilon = \pm 1$ (und zwar, wie man leicht sieht, $+1$). Daher hat die Abbildung l_p der ganzen S_m auf sich den Grad $2\epsilon = \pm 2$ (und zwar $+2$).

11. *Beweis des Satzes* IIb. Die stetigen Multiplikationen in Γ und Γ' seien durch

$$pq = l_p(q) = r_q(p) \qquad \text{bzw,} \qquad p'q' = l'_{p'}(q') = r'_{q'}(p')$$

gegeben; die zugehörigen Grade seien c_l, c_r, $c_{l'}$, $c_{r'}$; sie sind sämtlich $\neq 0$. Wir definieren in der Mannigfaltigkeit $\Gamma \times \Gamma'$, deren Punkte mit $p \times p'$, $q \times q'$, $\cdots$ bezeichnet werden, eine stetige Multiplikation durch die Festsetzung

$$(p \times p')\cdot(q \times q') = pq \times p'q' = L_{p\times p'}(q \times q') = R_{q\times q'}(p \times p');$$

die zugehörigen Grade seien C_L, C_R. Der Satz ist bewiesen, sobald gezeigt ist:

$$C_L = c_l \cdot c_{l'}, \qquad C_R = c_r \cdot c_{r'}.$$

Die Gültigkeit dieser Gleichheiten ist in dem folgenden *Hilfssatz* enthalten:

f und f' seien Abbildungen[20] der Mannigfaltigkeiten[2] A und A' in die Mannigfaltigkeiten B bzw. B', welche die gleichen Dimensionen haben wie A bzw. A'; die Grade von f und f' seien c bzw. c'. Dann hat die Abbildung F von $A \times A'$ in $B \times B'$, die durch

$$F(p \times p') = f(p) \times f'(p')$$

gegeben ist, wobei p, p' die Punkte von A bzw. A' durchlaufen, den Grad cc'.

Für den Beweis ersetzen wir f und f' durch so gute simpliziale Approximationen f_1, f'_1, daß auch diese die Grade c, c' haben, und daß auch die Abbildung F_1 von $A \times A'$ in $B \times B'$, die durch

$$F_1(p \times p') = f_1(p) \times f'_1(p')$$

gegeben ist, den gleichen Grad C hat wie F. Die Grundsimplexe der Zerlegungen von A, A', B, B', welche den simplizialen Abbildungen f_1, f'_1 zugrunde liegen, seien mit u_i, u'_j, v_k, v'_l bezeichnet; dann bilden die Produkte $u_i \times u'_j$ und $v_k \times v'_l$ die Grundzellen von Zellenzerlegungen der Mannigfaltigkeiten $A \times A'$ bzw. $B \times B'$. Durch F_1 wird jede Zelle $u_i \times u_j$ affin abgebildet, und zwar folgendermaßen: ist

$$f_1(u_i) = 0 \qquad \text{oder} \qquad f'_1(u'_j) = 0,$$

wird also die Dimension wenigstens eines der Simplexe u_i, u'_j durch die Abbildung f_1 oder f'_1 erniedrigt, so wird auch die Dimension der Zelle $u_i \times u'_j$ durch die Abbildung F_1 erniedrigt, es ist also $F_1(u_i \times u'_j) = 0$; ist

$$f_1(u_i) = \epsilon v_k, \qquad f'_1(u'_j) = \epsilon' v'_l, \qquad \epsilon = \pm 1, \qquad \epsilon' = \pm 1,$$

[20] Alle vorkommenden "Abbildungen" von Mannigfaltigkeiten sollen ***eindeutig und stetig*** sein.

sind also die beiden Abbildungen f_1, f_1' nicht-singulär, so ist auch die Abbildung F_1 von $u_i \times u_j'$ nicht-singulär, und es ist, wie sich aus bekannten Vorzeichenregeln bei der Bildung topologischer Produkte ergibt,

$$F_1(u_i \times u_j') = \epsilon\epsilon'(v_k \times v_l').$$

Jetzt lehrt eine leichte Abzählung: die algebraische Bedeckungszahl—d.h. die Anzahl der positiven Bedeckungen, vermindert um die Anzahl der negativen Bedeckungen—einer festen Grundzelle von $B \times B'$, etwa der Zelle $v_k \times v_l'$, also der Grad C von F_1, ist gleich dem Produkt der algebraischen Bedeckungszahlen von v_k und v_l' bei den Abbildungen f_1 bzw. f_1', also gleich cc'.

§2. Irreduzible Erzeugenden-Systeme und maximale Elemente eines Homologie-Ringes. Umformung des Satzes I.

12. *Vorbemerkungen.* Es sei M eine n-dimensionale Mannigfaltigkeit. Wie schon in Nr. 3 festgesetzt, bezeichnen wir die Dimension eines Elementes z von $\mathfrak{R}(M)$ mit $d(z)$ und verstehen unter seiner dualen Dimension die Zahl $\delta(z) = n - d(z)$.

Bekanntlich ist für homogen-dimensionale z, z' auch $z \cdot z'$ homogen-dimensional und

$$\delta(z \cdot z') = \delta(z) + \delta(z')\text{[21]} \tag{2.1}$$

sowie

$$z' \cdot z = \pm z \cdot z', \tag{2.2}$$

und zwar[22]

$$z' \cdot z = (-1)^{\delta(z) \cdot \delta(z')} z \cdot z', \tag{2.3}$$

also speziell

$$z \cdot z = 0 \qquad \text{bei ungeradem } \delta(z). \tag{2.4}$$

13. *Erzeugenden-Systeme.* Die homogen-n-dimensionalen Elemente von $\mathfrak{R}(M)$, also die rationalen Vielfachen der Eins des Ringes, nennen wir die "skalaren" Elemente von $\mathfrak{R}(M)$.

Unter einem "Erzeugenden-System" von $\mathfrak{R}(M)$ verstehen wir ein solches System von homogen-dimensionalen, nicht-skalaren Elementen $z_1, z_2, \cdots, z_L$, daß man alle erhält, wenn man auf $z_1, z_2, \cdots, z_L$ und 1 die Operationen der gegenseitigen Multiplikation, der Multiplikation mit rationalen Koeffizienten und der Addition ausübt.

Auf Grund der Regel (2.2) kann man jedes Element von $\mathfrak{R}(M)$ auf wenigstens eine Weise als Polynom in den z_λ, d.h. als Summe von Ausdrücken

$$t \cdot z_1^{\alpha_1} \cdot z_2^{\alpha_2} \cdot \cdots \cdot z_L^{\alpha_L}, \qquad \alpha_\lambda \geqq 0, \tag{2.5}$$

mit rationalen Koeffizienten t schreiben.

[21] Dem Null-Element des Ringes $\mathfrak{R}(M)$ wird *jede* Dimensionszahl zugeschrieben.

[22] Man vergl. z.B. Lefschetz, *Topology* [New York 1930], 166.

Ein Erzeugenden-System heißt "irreduzibel", wenn keines seiner echten Teilsysteme bereits ein Erzeugenden-System ist.

Offenbar ist in jedem Erzeugenden-System wenigstens ein irreduzibles Erzeugenden-System enthalten.

Man zeigt übrigens leicht, daß die Anzahl l der Elemente eines irreduziblen Erzeugenden-Systems von $\mathfrak{R}(M)$ nicht von diesem speziellen System abhängt, sondern eine Invariante von M ist; wir werden diese Tatsache, die wir vorläufig nicht benutzen, später beweisen (Nr. 31).

14. *Maximale Elemente.* Ein Element z von $\mathfrak{R}(M)$ heißt "maximal", wenn es 1) homogen-dimensional und nicht-skalar ist, und wenn es 2) nicht in dem Teilring von $\mathfrak{R}(M)$ enthalten ist, der von den homogen-dimensionalen Elementen z' von $\mathfrak{R}(M)$ mit $d(z') > d(z)$ erzeugt wird.[23]

Wir behaupten: *Jedes Element z_λ eines irreduziblen Erzeugenden-Systems $(z_1, z_2, \cdots, z_l)$ ist maximal.*

Beweis: Daß z_λ homogen-dimensional und nicht-skalar ist, ist in der Definition des Erzeugenden-Systems enthalten. Wäre z_λ nicht maximal, so wäre z_λ Element des Ringes $\mathfrak{U}$, der von allen homogen-dimensionalen Elementen z' mit $d(z') > d(z_\lambda)$ erzeugt wird. Nun läßt sich aber jedes dieser z' als Polynom in den Erzeugenden $z_1, z_2, \cdots, z_l$ schreiben, und hierbei tritt aus Dimensionsgründen das Element z_λ nicht auf; aus $z_\lambda \in \mathfrak{U}$ würde daher folgen, daß auch z_λ selbst ein Polynom in den von z_λ verschiedenen Elementen des Systems $(z_1, z_2, \cdots, z_l)$ wäre; dann würde aber dieses System, wenn man aus ihm z_λ wegließe, immer noch ein Erzeugenden-System bleiben—entgegen seiner Irreduzibilitäts-Eigenschaft.

15. *Umformung des Satzes I.*

Satz Ia. *$(z_1, z_2, \cdots, z_l)$ sei ein irreduzibles Erzeugenden-System des Ringes $\mathfrak{R}(\Gamma)$ einer Γ-Mannigfaltigkeit. Dann ist*

$$z_1 \cdot z_2 \cdot \cdots \cdot z_l \neq 0. \tag{2.6}$$

Satz Ib. *z sei ein maximales Element des Ringes $\mathfrak{R}(\Gamma)$ einer Γ-Mannigfaltigkeit. Dann ist $\delta(z)$ ungerade.*

Wir werden diese beiden Sätze im §4 beweisen. Jetzt wollen wir nur zeigen, daß aus ihnen der Satz I (Nr. 2, 3) folgt; dies wird geschehen sein, sobald wir bewiesen haben:

Es sei $(z_1, z_2, \cdots, z_l)$ ein irreduzibles Erzeugenden-System von $\mathfrak{R}(\Gamma)$, und es sei bekannt, daß die Sätze Ia und Ib gelten; dann haben $z_1, z_2, \cdots, z_l$ die in Nr. 3 genannten Eigenschaften.

Zunächst ergibt sich aus Nr. 14, daß alle Elemente z_i maximal, also aus Satz Ib, daß alle Zahlen

$$\delta(z_i) = m_i, \qquad i = 1, 2, \cdots, l, \tag{2.7}$$

[23] **Insbesondere ist jedes homogen $(n-1)$-dimensionale Element, das $\neq 0$ ist, maximal.**

ungerade sind. Nach (2.3) ist daher

$$z_j \cdot z_i = -z_i \cdot z_j , \tag{2.8}$$

also speziell

$$z_i \cdot z_i = 0. \tag{2.9}$$

Da die z_i ein Erzeugenden-System bilden, kann man jedes Element von $\mathfrak{R}(\Gamma)$ als Summe von Ausdrücken (2.5)—mit $L = l$—darstellen; infolge von (2.9) kann man sich dabei aber auf die Exponenten $\alpha_\lambda = 0$ und $\alpha_\lambda = 1$ beschränken; das heißt: jedes Element z von $\mathfrak{R}(\Gamma)$ läßt sich auf wenigstens eine Weise als lineare Verbindung mit rationalen Koeffizienten der Elemente

$$1;\quad z_i;\quad z_{i_1} \cdot z_{i_2}\ (i_2 < i_2);\quad z_{i_1} \cdot z_{i_2} \cdot z_{i_3}\ (i_1 < i_2 < i_3);\ \cdots;\quad z_1 \cdot z_2 \cdot \cdots \cdot z_l \tag{2.10}$$

darstellen, also in der Form

$$\begin{aligned} z = t + \sum t_i z_i + \sum t_{i_1 i_2} z_{i_1} \cdot z_{i_2} + \sum t_{i_1 i_2 i_3} z_{i_1} \cdot z_{i_2} \cdot z_{i_3} + \\ + \cdots + t_{12\ldots l} z_1 \cdot z_2 \cdot \cdots \cdot z_l , \end{aligned} \tag{2.11}$$

wobei die Koeffizienten t, t_i, $t_{i_1 i_2}$, $\cdots$ rational sind und die Indices die unter (2.10) angedeuteten Bedingungen erfüllen. Wir haben zu zeigen, daß die Elemente (2.10) eine additive Basis bilden, d.h. daß sie linear unabhängig sind (in Bezug auf rationale Koeffizienten), mit anderen Worten: in einer Darstellung (2.11) des Elementes $z = 0$ verschwinden alle Koeffizienten auf der rechten Seite.

Es sei also

$$\begin{aligned} 0 = t + \sum t_i z_i + \sum t_{i_1 i_2} z_{i_1} \cdot z_{i_2} + \sum t_{i_1 i_2 i_3} z_{i_1} \cdot z_{i_2} \cdot z_{i_3} + \\ + \cdots + t_{12\ldots l} z_1 \cdot z_2 \cdot \cdots \cdot z_l . \end{aligned} \tag{2.11$_0$}$$

Wir multiplizieren mit $z_1 \cdot z_2 \cdot \cdots \cdot z_l$; auf Grund des assoziativen Gesetzes und der Regeln (2.8), (2.9) verschwinden auf der rechten Seite alle Produkte, in denen ein Index zweimal auftritt, und es entsteht daher die Gleichung

$$0 = t \cdot z_1 \cdot z_2 \cdot \cdots \cdot z_l ;$$

nach Satz Ia folgt hieraus—da der Koeffizientenbereich ein Körper ist—:

$$t = 0.$$

Wir betrachten einen Index i und multiplizieren (2.11$_0$) mit dem Produkt aller der z_j, für welche $j \neq i$ ist; aus analogen Gründen wie soeben entsteht:

$$0 = t_i \cdot z_1 \cdot z_2 \cdot \cdots \cdot z_l ,$$

also folgt wie soeben:

$$t_i = 0.$$

Wir betrachten zwei Indizes i_1, i_2 und multiplizieren (2.11_0) mit dem Produkt aller der z_j, für welche $j \neq i_1$ und $j \neq i_2$ ist; es ergibt sich

$$t_{i_1 i_2} = 0.$$

So fortfahrend erkennt man, daß in der Tat alle Koeffizienten auf der rechten Seite von (2.11_0) gleich 0 sind.

Somit bilden die Elemente (2.10) eine Basis, und die Elemente von $\mathfrak{R}(\Gamma)$ lassen sich in eineindeutiger Weise mit den Ausdrücken (2.11) identifizieren. Daß für die Multiplikation die antikommutative Regel (2.8) gilt, wurde schon gezeigt. Für den vollständigen Beweis des Satzes I fehlt nur noch die Bestätigung der in Nr. 3 angegebenen Dimensions-Regeln.

Aus dem Satz Ia folgt, daß die Dimensionszahl $d(z_1 \cdot z_2 \cdot \cdots \cdot z_l) = d_0$ wohlbestimmt und $\geqq 0$ ist; für $i_1 < i_2 < \cdots < i_r$ ist $d(z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r}) \geqq d_0$, also ist infolge von (2.11) auch $d(z) \geqq d_0$ für jedes Element z von $\mathfrak{R}(\Gamma)$; da es 0-dimensionale Elemente z gibt, ist daher $d_0 = 0$. Dies ist, wenn Γ die Dimension n hat, gleichbedeutend mit $\delta(z_1 \cdot z_2 \cdot \cdots \cdot z_l) = n$; wenn wir $\delta(z_i) = m_i$ setzen, ist daher nach (2.1)

$$n = m_1 + m_2 + \cdots + m_l .$$

Allgemein ergibt sich für $i_1 < i_2 < \cdots < i_r$, wieder nach (2.1),

$$\begin{aligned} d(z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r}) &= n - \delta(z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r}) \\ &= n - m_{i_1} - m_{i_2} - \cdots - m_{i_r} . \end{aligned}$$

Daß schließlich alle m_i ungerade sind, wurde schon durch (2.7) festgestellt.

Damit ist der Satz I vollständig bewiesen—unter der Voraussetzung, daß die Sätze Ia und Ib gelten, die im §4 bewiesen werden sollen.

16. *Hilfssätze.*[24] Wir stellen hier noch einige einfache Tatsachen zusammen, die wir später (§4) brauchen werden.

M sei eine n-dimensionale Mannigfaltigkeit und $(z_1, z_2, \cdots, z_l)$ ein beliebiges Erzeugenden-System von $\mathfrak{R}(M)$. Unter $\mathfrak{U}$ verstehen wir die Menge aller derjenigen Elemente, die sich in der Gestalt

$$w_2 \cdot z_2 + w_3 \cdot z_3 + \cdots + w_l \cdot z_l$$

schreiben lassen, wobei die w_i beliebige Elemente von $\mathfrak{R}(M)$ sind.

Wir behaupten:

(a) *Jedes homogen-dimensionale Element z mit $d(z_1) < d(z) < n$ gehört zu $\mathfrak{U}$.*

(b) *Ist z_1 in $\mathfrak{U}$ enthalten, so ist das Erzeugenden-System $(z_1, z_2, \cdots, z_l)$ reduzibel.*

Beweis von (a): Man schreibe z als lineare Verbindung von Potenzprodukten $z_1^{\alpha_1} \cdot z_2^{\alpha_2} \cdot \cdots \cdot z_l^{\alpha_l}$ (mit rationalen Koeffizienten); da z und alle z_i homogen-

[24] **Die Hilfssätze der Nummern 16–18 und 20–22, die an und für sich kaum Interesse verdienen, werden erst in den Nummern 28 und 29 angewandt.**

dimensional sind, kann man sich dabei offenbar auf solche Potenzprodukte beschränken, die selbst die Dimension $d(z)$ haben; für ein solches Potenzprodukt ist dann nach (2.1)

$$\alpha_1 \cdot \delta(z_1) + \alpha_2 \cdot \delta(z_2) + \cdots + \alpha_l \cdot \delta(z_l) = \delta(z);$$

da $\delta(z) < \delta(z_1)$ ist, folgt hieraus $\alpha_1 = 0$, und da $\delta(z) > 0$ ist, folgt weiter, daß *nicht* $\alpha_2 = \cdots = \alpha_l = 0$ ist; dies bedeutet: jedes Potenzprodukt enthält wenigstens eines der Elemente $z_2, \cdots, z_l$ als Faktor, d.h. z gehört zu $\mathfrak{U}$.

Beweis von (b): Es sei $z_1 \in \mathfrak{U}$, also

$$z_1 = \sum_{j=2}^{l} w_j \cdot z_j; \tag{2.12}$$

dabei kann man, da die z_i homogen-dimensional sind, auch die Elemente w_j als homogen-dimensional und $d(w_j \cdot z_j) = d(z_1)$, also $\delta(w_j) + \delta(z_j) = \delta(z_1)$, annehmen. Da die $\delta(z_j) > 0$ sind (Nr. 13), sind daher alle $\delta(w_j) < \delta(z_1)$. Hieraus folgt, analog wie im Beweis von (a): stellt man w_j als lineare Verbindung von Potenzprodukten $z_1^{\alpha_1} \cdot z_2^{\alpha_2} \cdot \cdots \cdot z_l^{\alpha_l}$ dar, so ist immer $\alpha_1 = 0$; jedes w_j ist also durch $z_2, \cdots, z_l$ allein auszudrücken, und nach (2.12) ist dann auch z_1 durch $z_2, \cdots, z_l$ auszudrücken. Aber dann erzeugen bereits $z_2, \cdots, z_l$ den Ring $\mathfrak{R}(M)$.

17. Es sei z ein homogen-dimensionales Element von $\mathfrak{R}(M)$ und $d(z) < n$. Unter $\mathfrak{V}$ verstehen wir die Menge aller derjenigen Elemente, die sich als Summen von Produkten $w \cdot v$ schreiben lassen, wobei die Elemente v homogen-dimensional mit

$$d(v) \neq d(z), \qquad d(v) < n \tag{2.13}$$

und die Elemente w beliebig sind.

Wir behaupten:

(c) *Ist z in $\mathfrak{V}$ enthalten, so ist z nicht maximal.*

Beweis: Es sei z in $\mathfrak{V}$ enthalten, also

$$z = \sum w_h \cdot v_h; \tag{2.14}$$

hierin sind die v_h homogen-dimensional und erfüllen (2.13); da z homogen-dimensional ist, dürfen wir auch die w_h als homogen-dimensional und $d(w_h \cdot v_h) = d(z)$, also $\delta(w_h) + \delta(v_h) = \delta(z)$, für alle h annehmen. Dann ist $\delta(v_h) \leqq \delta(z)$, also, da nach (2.13) $\delta(v_h) \neq \delta(z)$ ist, $\delta(v_h) < \delta(z)$, $d(v_h) > d(z)$. Da nach (2.13) andererseits $\delta(v_h) > 0$ ist, ist auch $\delta(w_h) < \delta(z)$, $d(w_h) > d(z)$. Aus $d(v_h) > d(z)$ und $d(w_h) > d(z)$ folgt nach (2.14), daß z nicht maximal ist.

18. *Bemerkung über Ideale in $\mathfrak{R}(M)$.* Da die Multiplikation in dem Ring $\mathfrak{R}(M)$ im allgemeinen nicht kommutativ ist, hat man unter den Idealen des Ringes zwischen Links-, Rechts- und zweiseitigen Idealen zu unterscheiden. Jedoch gilt folgender Satz:

Es seien $x_1, x_2, \cdots, x_m$ *homogen-dimensionale* Elemente von $\mathfrak{R}(M)$ und $\mathfrak{X}$ das von ihnen erzeugte Links-Ideal, d.h. die Menge aller Elemente der Gestalt

$$w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_m \cdot x_m \tag{2.15}$$

mit willkürlichen Elementen w_i; dann ist $\mathfrak{X}$ zugleich Rechts-Ideal, also *zweiseitig.*

Zum Beweis ziehen wir eine volle additive Basis $(Z_1, Z_2, \cdots, Z_q)$ von $\mathfrak{R}(M)$ heran, deren Elemente Z_h wir als homogen-dimensional annehmen dürfen. Dann ist die Menge $\mathfrak{X}$ aller Elemente (2.15) identisch mit der Menge aller linearen Verbindungen der Produkte $Z_h \cdot x_i$ mit rationalen Koeffizienten ($h = 1, 2, \cdots, q$; $i = 1, 2, \cdots, m$). Nach (2.2) ist aber, da die Z_h und die x_i homogen-dimensional sind, $Z_h \cdot x_i = \pm x_i \cdot Z_h$, und daher ist $\mathfrak{X}$ auch die Menge aller linearen Verbindungen der Produkte $x_i \cdot Z_h$, also die Menge aller Elemente

$$x_1 \cdot w_1 + x_2 \cdot w_2 + \cdots + x_m \cdot w_m$$

mit willkürlichen w_i; diese Menge ist aber das von $x_1, x_2, \cdots, x_m$ erzeugte Rechts-Ideal.

Auf Grund dieses Satzes sind insbesondere die Mengen $\mathfrak{U}$ und $\mathfrak{B}$, die wir in Nr. 16 und 17 betrachtet haben, *zweiseitige Ideale.*

§3. Eigenschaften des Ringes $\mathfrak{R}(M \times M)$.

In Nr. 8 wurde die Rolle angedeutet, welche die Produkt-Mannigfaltigkeit $\Gamma \times \Gamma$ beim Beweise des Satzes I spielt. Der gegenwärtige Paragraph enthält lediglich eine, auf diesen Zweck zugeschnittene, Zusammenstellung und Formulierung bekannter Eigenschaften des Ringes $\mathfrak{R}(M \times M)$; dabei bezeichnet M eine beliebige Mannigfaltigkeit. Der Koeffizientenbereich ist wie immer der Körper der rationalen Zahlen.[25]

19. Zu je zwei Elementen x, y von $\mathfrak{R}(M)$ gehört ein Element $x \times y$ von $\mathfrak{R}(M \times M)$. Diese Produktbildung ist mit der Addition distributiv verknüpft. Sind x und y homogen-dimensional, so ist auch $x \times y$ homogen-dimensional und $d(x \times y) = d(x) + d(y)$.

Umgekehrt läßt sich jedes Element von $\mathfrak{R}(M \times M)$ als $\sum (x_h \times y_h)$ darstellen; genauer: ist das System $(Z_1, Z_2, \cdots, Z_q)$ eine volle additive Basis von $\mathfrak{R}(M)$—d.h. die Vereinigung von Homologiebasen aller Dimensionen—, so bilden die $Z_i \times Z_k$ eine volle additive Basis von $\mathfrak{R}(M \times M)$; die Elemente von $\mathfrak{R}(M \times M)$ sind also in eineindeutiger Weise als $\sum t_{ik}(Z_i \times Z_k)$ mit rationalen t_{ik} auszudrücken. Damit sind die additiven Eigenschaften von $\mathfrak{R}(M \times M)$ gegeben.

[25] Wegen der additiven Eigenschaften von $\mathfrak{R}(M \times M)$ vergl. man z.B. Alexandroff-Hopf, a.a.O.[6], Kap. VII, §3; jedoch hat man die dortigen Betrachtungen dadurch abzuändern (und wesentlich zu vereinfachen), daß man rationale Koeffizienten verwendet; wegen der multiplikativen Eigenschaften, insbesondere unserer Formel (3.1), vergl. man Lefschetz, a.a.O.[22], Chapter V, §3, insbesondere Formel (21).—Übrigens darf in unserem ganzen §3 der Koeffizientenbereich ein *beliebiger* Körper sein.

Die Multiplikation ist nun durch die Regel

$$(x \times y) \cdot (x' \times y') = (-1)^{\delta(x)\delta(y')}(x \cdot x' \times y \cdot y'), \tag{3.1}$$

welche *für homogen-dimensionale* x, y' gilt, und durch die distributiven Gesetze vollständig bestimmt.

20. *Hilfssätze.*[24] $\mathfrak{X}$ sei eine additive Untergruppe von $\mathfrak{R}(M)$, die rational abgeschlossen ist, d.h.: aus $x \in \mathfrak{X}$ folgt $tx \in \mathfrak{X}$ für jede rationale Zahl t. Ferner seien x, y, und zwar $y \neq 0$, zwei solche Elemente von $\mathfrak{R}(M)$, daß sich das Element $y \times x$ von $\mathfrak{R}(M \times M)$ in der Form

$$y \times x = \sum_h (y_h \times x_h) \tag{3.2}$$

darstellen läßt, wobei die x_h Elemente von $\mathfrak{X}$ sind, während von den Elementen y_h nichts vorausgesetzt wird.

Behauptung: *Dann ist* $x \in \mathfrak{X}$.

Beweis: $(Z_1, Z_2, \cdots, Z_q)$ sei eine volle additive Basis in $\mathfrak{R}(M)$. Es sei

$$x = \sum_i a_i Z_i, \qquad y = \sum_j b_j Z_j,$$

$$x_h = \sum_i c_{hi} Z_i, \qquad y_h = \sum_j d_{hj} Z_j$$

mit rationalen a, b, c, d; dann folgt aus (3.2), daß

$$b_j a_i = \sum_h d_{hj} c_{hi},$$

also

$$b_j x = \sum_h d_{hj} x_h \tag{3.3}$$

für jedes j ist. Da nach Voraussetzung $y \neq 0$ ist, ist wenigstens ein $b_j \neq 0$; es sei etwa $b_1 \neq 0$. Dann folgt aus (3.3)

$$x = \sum_h \frac{d_{h1}}{b_1} x_h \in \mathfrak{X}.$$

21. $\mathfrak{X}$ sei ein (zweiseitiges) Ideal in $\mathfrak{R}(M)$, welches von homogen-dimensionalen Elementen $x_1, x_2, \cdots, x_m$ erzeugt wird (man vergl. Nr. 18). Unter $\mathfrak{X}^*$ verstehen wir die Menge aller derjenigen Elemente von $\mathfrak{R}(M \times M)$, welche sich in der Gestalt

$$\sum_h (y_h \times x'_h) \tag{3.4}$$

schreiben lassen, wobei die x'_h Elemente von $\mathfrak{X}$, die y_h beliebige Elemente von $\mathfrak{R}(M)$ sind.

Behauptung: $\mathfrak{X}^*$ *ist ein zweiseitiges Ideal in* $\mathfrak{R}(M \times M)$.

Beweis: Es sei wieder $(Z_1, Z_2, \cdots, Z_q)$ eine volle additive Basis in $\mathfrak{R}(M)$; dann ist $\mathfrak{X}$ identisch mit der Menge aller linearen Verbindungen der Produkte $Z_j \cdot x_i$ (mit rationalen Koeffizienten) und $\mathfrak{X}^*$ daher identisch mit der Menge aller linearen Verbindungen der Elemente $Z_k \times Z_j \cdot x_i$ ($j, k = 1, 2, \cdots, q$; $i = 1, 2, \cdots, m$). Da wir die Z_k als homogen-dimensional annehmen dürfen, ist aber nach (3.1)

$$Z_k \times Z_j x_i = \pm(Z_k \times Z_j) \cdot (1 \times x_i);$$

die $Z_k \times Z_j$ bilden eine volle additive Basis; folglich ist $\mathfrak{X}^*$ die Menge aller Summen

$$\sum_i W_i \cdot (1 \times x_i),$$

wobei W_i willkürliche Elemente von $\mathfrak{R}(M \times M)$ sind; dies bedeutet aber: $\mathfrak{X}^*$ ist das Links-Ideal, das von den Elementen $1 \times x_i$, $i = 1, 2, \cdots, m$, erzeugt wird. Da die x_i homogen-dimensional sind, sind auch die Elemente $1 \times x_i$ homogen-dimensional, und nach Nr. 18 ist $\mathfrak{X}^*$ daher ein zweiseitiges Ideal.

22. Wir bringen den hiermit bewiesenen Satz in Verbindung mit dem Satz aus Nr. 20. Ein Ideal $\mathfrak{X}$ hat die in Nr. 20 genannte Eigenschaft der rationalen Abgeschlossenheit, da man ja die rationalen Zahlen als die skalaren Elemente—d.h. die rationalen Vielfachen des Eins-Elementes—von $\mathfrak{R}(M)$ auffassen kann. Daher ergibt sich aus Nr. 20 und 21 das folgende *Lemma*:

Die Ideale $\mathfrak{X}$ und $\mathfrak{X}^*$ sollen wie in Nr. 21 *erklärt sein; ferner seien x und y Element von $\mathfrak{R}(M)$, für welche*

$$y \neq 0$$

$$y \times x \equiv 0 \qquad \text{mod } \mathfrak{X}^*$$

gilt; dann ist

$$x \equiv 0 \qquad \text{mod } \mathfrak{X}.$$

23. *Die Homomorphismen* Λ und P. M sei eine n-dimensionale Mannigfaltigkeit und $(Z_1, Z_2, \cdots, Z_q)$ eine volle additive Basis von $\mathfrak{R}(M)$; wir dürfen annehmen, daß alle Z_j homogen-dimensional sind, daß $Z_1 = 1$ und $d(Z_j) < n$ für $j > 1$ ist. Da die $Z_j \times Z_k$ eine Basis von $\mathfrak{R}(M \times M)$ bilden, besitzt jedes Element Q von $\mathfrak{R}(M \times M)$ eine und nur eine Darstellung

$$Q = \sum_{j,k} t_{jk}(Z_j \times Z_k) \tag{3.4}$$

mit rationalen t_{jk}. Wir setzen

$$\Lambda(Q) = \sum_k t_{1k} Z_k \tag{3.5}$$

Zum Beispiel ist für jedes Element y von $\mathfrak{R}(M)$

$$\Lambda(1 \times y) = y \tag{3.6}$$

und, falls $d(x) < n$ ist,

$$\Lambda(x \times y) = 0; \tag{3.7}$$

man bestätigt (3.6) und (3.7) einfach, indem man y, bzw. x und y, als lineare Verbindungen der Z_k schreibt.

Wir fassen Λ als Abbildung des Ringes $\mathfrak{R}(M \times M)$ in den Ring $\mathfrak{R}(M)$ auf und behaupten: Λ *ist eine Homomorphie.*

Daß Λ additiv homomorph, d.h. daß

$$\Lambda(Q_1 + Q_2) = \Lambda(Q_1) + \Lambda(Q_2)$$

ist, liest man unmittelbar aus (3.4) und (3.5) ab. Für den Beweis der multiplikativen Homomorphie, d.h. der Gültigkeit der Gleichheit

$$\Lambda(Q_1 \cdot Q_2) = \Lambda(Q_1) \cdot \Lambda(Q_2) \tag{3.8}$$

darf man sich infolge der Distributivität und der schon konstatierten additiven Homomorphie auf den Fall beschränken, daß Q_1, Q_2 Elemente der Basis $Z_j \times Z_k$ von $\mathfrak{R}(M \times M)$ sind. Es sei also

$$Q_1 = Z_h \times Z_i, \qquad Q_2 = Z_j \times Z_k.$$

Ist $h = j = 1$, so ist einerseits

$$\Lambda(Q_1) = Z_i, \qquad \Lambda(Q_2) = Z_k,$$

andererseits nach (3.1)

$$Q_1 \cdot Q_2 = 1 \times Z_i \cdot Z_k,$$

also nach (3.6)

$$\Lambda(Q_1 \cdot Q_2) = Z_i \cdot Z_k;$$

folglich gilt (3.8). Ist nicht $h = j = 1$, so ist wenigstens eines der Elemente $\Lambda(Q_1)$, $\Lambda(Q_2)$ gleich Null, also ist die rechte Seite von (3.8) Null; andererseits ist nach (3.1)

$$Q_1 \cdot Q_2 = \pm Z_h \cdot Z_j \times Z_i \cdot Z_k,$$

und hierin ist $d(Z_h \cdot Z_j) < n$; daher ist nach (3.7) auch die linke Seite $\Lambda(Q_1 \cdot Q_2)$ von (3.8) gleich Null.—Damit ist die Homomorphie-Eigenschaft von Λ bewiesen.

Setzt man im Anschluß an (3.4)

$$\mathsf{P}(Q) = \sum_j t_{j1} Z_j, \tag{3.9}$$

so ergibt sich ganz analog: P ist eine homomorphe Abbildung von $\mathfrak{R}(M \times M)$ in $\mathfrak{R}(M)$.

24. Aus (3.4), (3.5), (3.9) liest man ab, daß für jedes Element Q von $\mathfrak{R}(M \times M)$ Gleichungen

$$\begin{aligned} Q &= (1 \times \Lambda(Q)) + \sum_j (Z_j \times Y_j), \qquad d(Z_j) < n, \\ Q &= (\mathsf{P}(Q) \times 1) + \sum_k (X_k \times Z_k), \qquad d(Z_k) < n \end{aligned} \tag{3.10}$$

gelten.

Ferner ist klar: ist Q homogen $(n + r)$-dimensional, so sind $\Lambda(Q)$ und $\mathsf{P}(Q)$ homogen r-dimensional.

Es sei jetzt Q homogen $(n + r)$-dimensional und $r < n$. Dann ist in (3.4) $t_{11} = 0$, und daher erhält (3.4) mit Hilfe von (3.5) und (3.9) die Gestalt

$$Q = (1 \times \Lambda(Q)) + (\mathsf{P}(Q) \times 1) + \sum_{j=2}^{q} \sum_{k=2}^{q} t_{jk}(Z_j \times Z_k). \tag{3.11}$$

Da die Z_i homogen-dimensional sind, sind auch die $Z_j \times Z_k$ homogen-dimensional, und es ist $d(Z_j \times Z_k) = d(Z_j) + d(Z_k)$; daher sind in (3.11) nur solche $t_{jk} \neq 0$, für welche

$$d(Z_j) + d(Z_k) = n + r$$

ist; ferner ist in (3.11), da $j > 0$, $k > 0$ ist, immer

$$d(Z_j) < n, \qquad d(Z_k) < n$$

und folglich

$$d(Z_k) > r, \qquad d(Z_j) > r.$$

Mithin läßt sich (3.11) auch so ausdrücken:

$$\begin{cases} Q = (1 \times \Lambda(Q)) + (\mathsf{P}(Q) \times 1) + \sum\limits_{h} (x_h \times y_h), \\ x_h,\ y_h \text{ homogen-dimensional mit} \\ r < d(x_h) < n, \qquad r < d(y_h) < n. \end{cases} \tag{3.12}$$

Wir fassen zusammen: *Es gibt zwei solche Homomorphismen* Λ *und* P *des Ringes* $\mathfrak{R}(M \times M)$ *in den Ring* $\mathfrak{R}(M)$, *daß jedes Element* Q *von* $\mathfrak{R}(M \times M)$ *Darstellungen* (3.10) *und daß jedes homogen* $(n + r)$*-dimensionale Element* Q *mit* $r < n$ *eine Darstellung* (3.12) *besitzt.*

Es ist übrigens leicht zu sehen, daß die Abbildungen Λ und P, die wir durch (3.5) und (3.9) unter Benutzung einer speziellen Basis $\{Z_i\}$ erklärt haben, von der Wahl dieser Basis unabhängig sind.

§4. Beweis des Satzes I.

25. *Der Umkehrungs-Homomorphismus.* M und M' seien beliebige Mannigfaltigkeiten,[2] und F sei eine Abbildung von M in M'. Dann bewirkt F eine Abbildung des Ringes $\mathfrak{R}(M)$ in den Ring $\mathfrak{R}(M')$; wir bezeichnen auch diese Ring-Abbildung mit F; sie ist übrigens additiv homomorph, jedoch im allgemeinen nicht multiplikativ homomorph.

Es gilt der Satz:[26]

[26] H. Hopf, *Zur Algebra der Abbildungen von Mannigfaltigkeiten* [Journ. f.d.r.u.a. Math. **163** (1930), 71–88].—*Neue Begründung und Verallgemeinerung des Umkehrungs-Homomorphismus*: H. Freudenthal, *Zum Hopfschen Umkehrhomomorphismus* [Annals of Math. (2) **38** (1937), 847–853]; ferner: A. Komatu, *Über die Ringdualität eines Kompaktums* [Tôhoku Math. Journ. **43** (1937), 414–420]; H. Whitney, *On products in a complex* [Annals of Math. (2) **39** (1938), 397–432], (Theorem 6).—Die Eigenschaft 3 unseres Textes ist in meiner zitierten Arbeit nicht hervorgehoben, da dort nur gleichdimensionale Mannigfaltigkeiten betrachtet werden; sie ergibt sich aber unmittelbar aus jeder einzelnen der verschiedenen, in den soeben genannten Arbeiten enthaltenen, Definitionen von Φ; überdies ist sie eine Folge der Eigenschaft 2; hierzu vergl. man Nr. 11 meiner Arbeit "*Ein topologischer Beitrag zur reellen Algebra*" [Com. Math. Helvet.; erscheint nächstens].

Es existiert eine Abbildung Φ des Ringes $\mathfrak{R}(M')$ in den Ring $\mathfrak{R}(M)$ mit den folgenden drei Eigenschaften:

1) *Φ ist ein additiver und multiplikativer Homomorphismus;*

2) *Φ ist mit F durch die Funktionalgleichung*

$$F(\Phi(z)\cdot x) = z\cdot F(x) \tag{4.1}$$

verknüpft, in welcher x ein beliebiges Element von $\mathfrak{R}(M)$ und z ein beliebiges Element von $\mathfrak{R}(M')$ ist;

3) *ist z homogen-dimensional, so ist auch $\Phi(z)$ homogen-dimensional, und zwar ist $\delta(\Phi(z)) = \delta(z)$*, also

$$d(\Phi(z)) = d(z) + d(M) - d(M').$$

Φ heißt der "Umkehrungs-Homomorphismus" von F.

26. *Ansatz zum Beweis der Sätze* Ia *und* Ib. In der n-dimensionalen Mannigfaltigkeit Γ sei eine stetige Multiplikation gegeben (Nr. 1). Da wir die Punktepaare (p, q) von Γ als die Punkte $p \times q$ der Produkt-Mannigfaltigkeit $\Gamma \times \Gamma$ deuten können, ist die stetige Multiplikation gleichbedeutend mit einer Abbildung F von $\Gamma \times \Gamma$ in Γ; diese ist durch $F(p \times q) = pq$ bestimmt.

Wie in Nr. 25 bezeichnen wir auch die durch F bewirkte Abbildung des Ringes $\mathfrak{R}(\Gamma \times \Gamma)$ in den Ring $\mathfrak{R}(\Gamma)$ mit F. Das Element von $\mathfrak{R}(\Gamma)$, das durch einen einfach gezählten Punkt repräsentiert wird, heiße p. Das mit einer rationalen Zahl c multiplizierte Eins-Element von $\mathfrak{R}(\Gamma)$ bezeichnen wir kurz mit c. Dann sind die Grade c_l und c_r, die in Nr. 1 erklärt worden sind, offenbar durch

$$F(p \times 1) = c_l, \qquad F(1 \times p) = c_r \tag{4.2}$$

charakterisiert.

Φ sei der Umkehrungs-Homomorphismus von F. Dann folgt aus (4.1) und (4.2)

$$F(\Phi(z)\cdot(p \times 1)) = c_l z, \tag{4.3a}$$

$$F(\Phi(z)\cdot(1 \times p)) = c_r z \tag{4.3b}$$

für jedes Element z von $\mathfrak{R}(\Gamma)$.

Die Homomorphismen $\mathsf{\Lambda}$ und P sind in Nr. 23, 24 erklärt worden; wir setzen

$$\mathsf{\Lambda}\Phi(z) = \lambda(z), \qquad \mathsf{P}\Phi(z) = \rho(z)$$

für jedes Element z von $\mathfrak{R}(\Gamma)$; dann sind λ und ρ *Homomorphismen des Ringes* $\mathfrak{R}(\Gamma)$ *in sich*, und nach Nr. 24, (3.10), gelten für jedes Element z von $\mathfrak{R}(\Gamma)$ Gleichungen

$$\Phi(z) = (1 \times \lambda(z)) + \sum_j (Z_j \times Y_j), \qquad d(Z_j) < n, \tag{4.4a}$$

$$\Phi(z) = (\rho(z) \times 1) + \sum_k (X_k \times Z_k), \qquad d(Z_k) < n. \tag{4.4b}$$

Nach Nr. 25, 3), ordnet Φ, da $d(\Gamma \times \Gamma) = 2n$, $d(\Gamma) = n$ ist, jedem homogen r-dimensionalen Element von $\mathfrak{R}(\Gamma)$ ein homogen $(n + r)$-dimensionales Element von $\mathfrak{R}(\Gamma \times \Gamma)$ zu; nach Nr. 24 ordnen Λ und P den homogen $(n + r)$-dimensionalen Elementen von $\mathfrak{R}(\Gamma \times \Gamma)$ homogen r-dimensionale Elemente von $\mathfrak{R}(\Gamma)$ zu; folglich sind λ und ρ *dimensionstreu.*

27. Jetzt sei

$$c_l \neq 0, \qquad c_r \neq 0, \tag{4.5}$$

also Γ eine Γ-Mannigfaltigkeit. Wir behaupten: *dann sind λ und ρ Automorphismen*[27] *von* $\mathfrak{R}(\Gamma)$.

Beweis: Ist $\lambda(z) = 0$, so ist nach (4.4a)

$$\Phi(z) = \sum (Z_i \times Y_i) \text{ mit } d(Z_i) < n;$$

aus $d(Z_i) < n$ folgt $Z_i \cdot p = 0$; folglich ist, nach (3.1),

$$\Phi(z) \cdot (p \times 1) = 0,$$

also nach (4.3a)

$$c_l z = 0,$$

also nach (4.5), da der Koeffizientenbereich ein Körper ist,

$$z = 0.$$

Das bedeutet: der Homomorphismus λ ist eineindeutig. Folglich ist die Determinante der Substitution, welche durch λ auf eine volle additive Basis von $\mathfrak{R}(\Gamma)$ ausgeübt wird, nicht Null. Da der Koeffizientenbereich ein Körper ist, folgt hieraus: λ ist eine Abbildung von $\mathfrak{R}(\Gamma)$ *auf* sich. Somit ist λ ein Automorphismus.– Analog beweist man die Behauptung für ρ.

Wir fassen zusammen, indem wir noch das Ergebnis von Nr. 24 heranziehen: *Für jede n-dimensionale Γ-Mannigfaltigkeit Γ sind zwei dimensionstreue Automorphismen λ und ρ von $\mathfrak{R}(\Gamma)$ und ein Homomorphismus Φ von $\mathfrak{R}(\Gamma)$ in $\mathfrak{R}(\Gamma \times \Gamma)$ ausgezeichnet; ist z ein homogen r-dimensionales Element von $\mathfrak{R}(\Gamma)$ und $r < n$, so gilt*

$$\begin{cases} \Phi(z) = (1 \times \lambda(z)) + (\rho(z) \times 1) + \sum (x_h \times y_h), \\ x_h,\, y_h \text{ homogen-dimensional mit} \\ \qquad r < d(x_h) < n, \qquad r < d(y_h) < n. \end{cases} \tag{4.6}$$

In diesem Satz sind alle Eigenschaften der Γ-Mannigfaltigkeiten enthalten, die wir im Folgenden benutzen werden.

28. *Beweis des Satzes* Ia (Nr. 15). Es sei $(z_1, z_2, \cdots, z_l)$ ein irreduzibles Erzeugenden-System für die n-dimensionale Γ-Mannigfaltigkeit Γ. Wir werden

[27] Unter einem "Automorphismus" eines Ringes $\mathfrak{R}$ ist ein Auto-*Isomorphismus* von $\mathfrak{R}$ *auf* sich zu verstehen.

durch vollständige Induktion nach k beweisen: das Produkt von je k voneinander verschiedenen Elementen dieses Systems ist $\neq 0$. Für $k = 1$ ist dies richtig, da ein irreduzibles Erzeugenden-System niemals die Null enthalten kann. Es sei für $k - 1$ bewiesen, und k Elemente des Systems, etwa $z_1, z_2, \cdots, z_k$, seien vorgelegt; da ihre Anordnung lediglich das Vorzeichen ihres Produktes beinflussen kann, dürfen wir annehmen, daß

$$d(z_1) \leqq d(z_j), \qquad j = 1, 2, \cdots, k \tag{4.7}$$

ist; nach Induktions-Annahme ist

$$z_2 \cdot z_3 \cdot \cdots \cdot z_k \neq 0,$$

also, da ρ ein Automorphismus ist, auch

$$\rho(z_2 \cdot z_3 \cdot \cdots \cdot z_k) \neq 0. \tag{4.8}$$

Wir setzen $\lambda(z_i) = z_i'$ für $i = 1, 2, \cdots, l$. Da λ ein Automorphismus ist, bilden auch $z_1', z_2', \cdots, z_l'$ ein irreduzibles Erzeugenden-System. Da λ dimensionstreu ist, ist

$$d(z_i') = d(z_i), \qquad i = 1, 2, \cdots, l \tag{4.9}$$

und nach (4.7)

$$d(z_1') \leqq d(z_j'), \qquad j = 1, 2, \cdots, k. \tag{4.7'}$$

Wir verstehen nun wie in Nr. 16—mit dem Unterschied, daß wir die dortigen z_i durch unsere neuen z_i' ersetzen—unter $\mathfrak{U}$ das Ideal in $\mathfrak{R}(\Gamma)$, das von $z_2', \cdots, z_l'$ erzeugt wird (man vergl. Nr. 18); ferner verstehen wir unter $\mathfrak{U}^*$ das, analog wie in Nr. 21 erklärte, zu $\mathfrak{U}$ gehörige Ideal in $\mathfrak{R}(\Gamma \times \Gamma)$, also die Menge derjenigen Elemente von $\mathfrak{R}(\Gamma \times \Gamma)$, die sich in der Gestalt $\sum (w_h \times u_h')$ mit $u_h' \in \mathfrak{U}$ schreiben lassen.

Schreiben wir für $j = 1, 2, \cdots, k$ das Element $\Phi(z_j)$ in der Form (4.6):

$$\Phi(z_j) = (1 \times z_j') + (\rho(z_j) \times 1) + \sum (x_h \times y_h), \tag{4.6$_j$}$$

so lautet auf Grund von (4.9) und (4.7′) die eine der Dimensions-Bedingungen aus (4.6)

$$d(z_1') \leqq d(z_j') < d(y_h) < n;$$

folglich gehören nach Nr. 16 (a) alle in den Gleichungen (4.6$_j$) auftretenden y_h zu $\mathfrak{U}$ und daher die Summe $\sum (x_h \times y_h)$ zu $\mathfrak{U}^*$; für $1 < j \leqq k$ gehört außerdem z_j' zu $\mathfrak{U}$, also $1 \times z_j'$ zu $\mathfrak{U}^*$. Daher sind in den Gleichungen (4.6$_j$) die folgenden Kongruenzen modulo $\mathfrak{U}^*$ enthalten:

$$\Phi(z_1) \equiv (1 \times z_1') + (\rho(z_1) \times 1),$$

$$\Phi(z_j) \equiv \qquad (\rho(z_j) \times 1), \qquad j = 2, \cdots, k.$$

Da $\mathfrak{U}^*$ ein zweiseitiges Ideal ist, dürfen wir diese Kongruenzen miteinander multiplizieren, und da Φ und ρ Homomorphismen sind, ergibt sich dabei, bei Beachtung von (3.1):

$$\Phi(z_1 \cdot z_2 \cdot \cdots \cdot z_k) \equiv (\rho(z_2 \cdot \cdots \cdot z_k) \times z_1') + (\rho(z_1 \cdot z_2 \cdot \cdots \cdot z_k) \times 1).$$

Wäre nun $z_1 \cdot z_2 \cdot \cdots \cdot z_k = 0$, so würde hieraus

$$\rho(z_2 \cdot \cdots \cdot z_k) \times z_1' \equiv 0 \bmod \mathfrak{U}^*$$

folgen; nach (4.8) und dem Lemma von Nr. 22 wäre dann

$$z_1' \equiv 0 \bmod \mathfrak{U},$$

also z_1' in $\mathfrak{U}$ enthalten; nach Nr. 16 (b) ist dies mit der Irreduzibilität des Erzeugenden-Systems $(z_1', z_2', \cdots, z_l')$ nicht verträglich. Folglich ist

$$z_1 \cdot z_2 \cdot \cdots \cdot z_k \neq 0,$$

was zu beweisen war.[28]

29. *Beweis des Satzes* Ib (Nr. 15). Im Ring der n-dimensionalen Γ-Mannigfaltigkeit Γ sei z ein homogen-dimensionales Element, für welches $d(z) < n$ und $\delta(z)$ *gerade* ist; wir haben zu zeigen, daß dann z nicht maximal ist.

Das Ideal $\mathfrak{V}$ sei wörtlich wie in Nr. 17 definiert; $\mathfrak{V}^*$ sei das Ideal in $\mathfrak{R}(\Gamma \times \Gamma)$, das analog wie in Nr. 21 durch $\mathfrak{V}$ bestimmt ist: es besteht aus allen Elementen $\sum (w_h \times v_h)$ mit $v_h \in \mathfrak{V}$.

In der Gleichung (4.6) für unser Element z gehören alle Elemente y_h zu $\mathfrak{V}$ und somit gehört die Summe $\sum (x_h \times y_h)$ zu $\mathfrak{V}^*$. Es gilt also die Kongruenz

$$\Phi(z) \equiv (1 \times \lambda(z)) + (\rho(z) \times 1) \qquad \bmod \mathfrak{V}^*. \tag{4.10}$$

Nach (3.1) ist

$$(1 \times \lambda(z)) \cdot (\rho(z) \times 1) = \rho(z) \times \lambda(z) \tag{4.11}$$

und, da $\delta(z)$ gerade ist, auch

$$(\rho(z) \times 1) \cdot (1 \times \lambda(z)) = \rho(z) \times \lambda(z). \tag{4.11'}$$

Aus (4.11) und (4.11') ergibt sich, daß man die rechte Seite von (4.10) nach der binomischen Formel potenzieren kann; tut man dies und beachtet man, daß Φ, ρ, λ Homomorphismen sind, so erhält man für jeden positiven Exponenten m:

$$\Phi(z^m) \equiv \sum_{r=0}^{m} \binom{m}{r} (\rho(z^{m-r}) \times \lambda(z^r)) \qquad \bmod \mathfrak{V}^*. \tag{4.12}$$

Nun ist, da $d(z) < n$ ist, für $r > 1$ immer $d(z^r) < d(z)$, also, da λ dimensionstreu ist, auch $d(\lambda(z^r)) < d(z)$ und daher $\lambda(z^r) \in \mathfrak{V}$ und

$$\rho(z^{m-r}) \times \lambda(z^r) \in \mathfrak{V}^* \qquad \text{für } r > 1;$$

[28] Der Beweis zeigt, daß für den Satz Ia der Koeffizientenbereich ein *beliebiger Körper* sein darf.

daher reduziert sich die Kongruenz (4.12) auf

$$\Phi(z^m) \equiv m(\rho(z^{m-1}) \times \lambda(z)) + (\rho(z^m) \times 1) \bmod \mathfrak{B}^*. \tag{4.13}$$

Dies gilt für jeden positiven Exponenten m. Da aus Dimensions-Gründen $z^m = 0$ für hinreichend große m ist, kann man m so wählen, daß

$$z^{m-1} \neq 0, \qquad z^m = 0$$

ist. Dann entsteht aus (4.13) die Kongruenz

$$m(\rho(z^{m-1}) \times \lambda(z)) \equiv 0 \bmod \mathfrak{B}^*,$$

also, da der Koeffizientenbereich der Körper der rationalen Zahlen ist,[29]

$$\rho(z^{m-1}) \times \lambda(z) \equiv 0 \bmod \mathfrak{B}^*. \tag{4.14}$$

Da $z^{m-1} \neq 0$ und ρ ein Automorphismus ist, ist auch $\rho(z^{m-1}) \neq 0$; nach dem Lemma in Nr. 22 folgt daher aus (4.14)

$$\lambda(z) \equiv 0 \bmod \mathfrak{B},$$

d.h. $\lambda(z) \in \mathfrak{B}$. Der Automorphismus λ ist dimensionstreu, und die Menge $\mathfrak{B}$ ist durch Dimensions-Eigenschaften definiert; daher geht $\mathfrak{B}$ bei Ausübung von λ^{-1} in sich über, und mit $\lambda(z)$ ist daher auch z in $\mathfrak{B}$ enthalten. Nach dem Hilfssatz Nr. 17 (c) ist mithin z nicht maximal —was zu beweisen war.

Mit den Sätzen Ia und Ib ist der Satz I vollständig bewiesen (Nr. 15).

§5. Maximale und minimale Elemente. Beweis des Satzes III

30. *Die Ränge l_s.* Wir knüpfen an Nr. 13 und 14 an. Es sei wieder M eine beliebige n-dimensionale Mannigfaltigkeit. Die additive Gruppe der homogen r-dimensionalen Elemente von $\mathfrak{R}(M)$, also die r-te Bettische Gruppe (in bezug auf rationale Koeffizienten) heiße $\mathfrak{B}_r$; ihr Rang p_r ist die r-te Bettische Zahl von M.

Für $0 \leqq r < n$ verstehen wir unter $\mathfrak{U}_r$ die Gruppe derjenigen Elemente von $\mathfrak{B}_r$, welche sich aus den Elementen der Gruppe

$$\mathfrak{B}_n + \mathfrak{B}_{n-1} + \cdots + \mathfrak{B}_{r+1}$$

durch Multiplikation und Addition erzeugen lassen, mit anderen Worten: welche nicht maximal sind. $\mathfrak{U}_r$ läßt sich auch als die Gesamtheit derjenigen Elemente charakterisieren, die sich in der Form $\sum_i x_i \cdot y_i$ schreiben lassen, wobei x_i, y_i homogen-dimensional mit

$$\delta(x_i) + \delta(y_i) = n - r, \qquad 0 < \delta(x_i), \qquad 0 < \delta(y_i)$$

sind.

Den Rang von $\mathfrak{U}_r$ nennen wir q_r, und wir setzen

$$p_{n-s} - q_{n-s} = l_s, \qquad s = 1, 2, \cdots, n.$$

[29] **Dies ist die einzige Stelle in der ganzen Arbeit, an welcher benutzt wird, daß der Koeffizientenkörper die *Charakteristik* 0 hat.**

Die Zahl l_s ist der Rang der Restklassengruppe $\mathfrak{B}_{n-s} - \mathfrak{U}_{n-s}$ und hat daher die folgende Bedeutung: es gibt ein System von l_s derartigen maximalen Elementen $y_1, y_2, \cdots, y_{l_s}$ mit $\delta(y_i) = s$, daß auch jede lineare Verbindung

$$u_1y_1 + u_2y_2 + \cdots + u_{l_s}y_{l_s}$$

mit rationalen Koeffizienten u_i, abgesehen von derjenigen mit $u_1 = \cdots = u_{l_s} = 0$, selbst maximal ist; dagegen gibt es für kein $l' > l_s$ ein System von l' derartigen Elementen.

31. Wir behaupten erstens: In einem Erzeugenden-System $(z_1, \cdots, z_L)$ von $\mathfrak{R}(M)$ ist die Anzahl der z_j mit $\delta(z_j) = s$ stets $\geqq l_s$.

Beweis: Es seien $y_1, \cdots, y_{l_s}$ maximale Elemente mit der soeben genannten Eigenschaft. Stellt man die y_i als Polynome in den Erzeugenden $z_1, \cdots, z_L$ dar, so haben diese Darstellungen aus Dimensions-Gründen die Gestalt

$$y_i = t_{i1}z_1 + \cdots + t_{im}z_m + Y_i ;$$

dabei sind $z_1, \cdots, z_m$ diejenigen Erzeugenden z_j, für welche $\delta(z_j) = s$ ist $(m \geqq 0)$, t_{ij} rationale Zahlen und Y_i Polynome in den z_k mit $d(z_k) > n - s$. Wäre $m < l_s$, so besäße das Gleichungssystem

$$\sum_{i=1}^{l_s} u_i t_{ij} = 0, \qquad j = 1, \cdots, m,$$

eine rationale Lösung $(u_1, \cdots, u_{l_s}) \neq (0, \cdots, 0)$, und es wäre

$$u_1y_1 + \cdots + u_{l_s}y_{l_s} = u_1Y_1 + \cdots + u_{l_s}Y_{l_s} ;$$

dieses Element wäre, wie die rechte Seite zeigt, nicht maximal — entgegen der Voraussetzung über die y_i.

Wir behaupten zweitens: Ist—bei Benutzung derselben Bezeichnungen wie soeben—$m > l_s$, so ist das Erzeugenden-System $(z_1, \cdots, z_L)$ reduzibel.

Beweis: Da $l' = m > l_s$ ist, kann man, wie am Schluß von Nr. 30 festgestellt wurde, Zahlen $u_1, \cdots, u_m$ so bestimmen, daß das Element

$$Z = u_1z_1 + \cdots + u_mz_m$$

nicht maximal ist, und daß nicht alle $u_j = 0$ sind; dann ist Z ein Polynom in den z_k mit $d(z_k) > n - s$, also erst recht in den z_i mit $i > m$, und man kann, wenn etwa $u_1 \neq 0$ ist, z_1 durch die z_i mit $i > 1$ ausdrücken und somit das Erzeugenden-System reduzieren.

Damit haben wir festgestellt: *In jedem irreduziblen Erzeugenden-System $(z_1, \cdots, z_l)$ ist die Anzahl derjenigen z_i, für welche $\delta(z_i) = s$ ist, gleich l_s; die Anzahl aller Erzeugenden eines irreduziblen Systems ist daher immer*

$$l = l_1 + \cdots + l_n .$$[30]

[30] Hierdurch ist jeder Mannigfaltigkeit M eine Invariante l zugeordnet, welche etwa der "Rang" von M heißen möge. Für eine Γ-Mannigfaltigkeit ist, wie sich aus dem Satz (b) in Nr. 4 und seiner Herleitung ergibt, die Summe der Bettischen Zahlen gleich 2^l; andererseits ist für eine Gruppen-Mannigfaltigkeit, wie Cartan mit Hilfe der Integral-Invarianten berechnet hat[8], (b), 24, der hier auftretende Exponent gleich dem "Rang" der *Gruppe*,

32. *Die minimalen Elemente.* Es sei v ein homogen-dimensionales Element von $\mathfrak{R}(M)$ und $d(v) > 0$. Wir betrachten seine Vielfachen, d.h. die Elemente $x \cdot v$, wobei x die Elemente von $\mathfrak{R}(M)$ durchläuft; unter diesen Vielfachen sind außer dem Null-Element erstens alle rationalen Vielfachen von v enthalten—das sind Elemente der Dimension $d(v)$—und zweitens, falls nur $v \neq 0$ ist, nach dem Poincaré-Veblenschen Dualitäts-Satz alle Elemente der Dimension 0; für alle Vielfachen ist $d(x \cdot v) \leqq d(v)$. Wir definieren nun: das Element v heißt "minimal", wenn es keine Vielfachen $x \cdot v$ besitzt, für welche

$$x \cdot v \neq 0, \qquad 0 < d(x \cdot v) < d(v)$$

ist. Diese Definition ist offenbar gleichwertig mit der folgenden: v ist minimal, wenn für alle homogen-dimensionalen x, für welche

$$0 < \delta(x) < d(v), \text{ also } n - d(v) < d(x) < n,$$

ist, die Produkte $x \cdot v = 0$ sind.

Dabei ist, wie oben gesagt, $d(v) > 0$ vorausgesetzt; diese Verabredung ist analog der früher getroffenen, die n-dimensionalen Elemente nicht als maximal zu bezeichnen. Dagegen ist das Element 0 des Ringes $\mathfrak{R}(M)$ minimal.[31]

33. *Ein Dualitätssatz.* Die Gruppen $\mathfrak{U}_r$ sind in Nr. 30 erklärt worden. Wir behaupten:

Das homogen s-dimensionale Element v ist dann und nur dann minimal, wenn es ein Annullator der Gruppe $\mathfrak{U}_{n-s}$ ist, d.h. wenn $u \cdot v = 0$ für jedes Element u aus $\mathfrak{U}_{n-s}$ gilt $(s = 1, \cdots, n)$.

Beweis: Es sei erstens v minimal und $u \in \mathfrak{U}_{n-s}$; dann ist $u = \sum x_i \cdot y_i$, wobei die x_i, y_i homogen-dimensional sind, mit

$$\delta(x_i) + \delta(y_i) = s, \qquad \delta(x_i) > 0, \qquad \delta(y_i) > 0,$$

also

$$0 < \delta(y_i) < s = d(v);$$

hieraus folgt $y_i \cdot v = 0$ für jedes i, also $u \cdot v = 0$. Es sei zweitens v homogen s-dimensional, aber nicht minimal; dann gibt es ein homogen-dimensionales y mit $0 < \delta(y) < s$ und $y \cdot v \neq 0$, und nach dem Poincaré-Veblenschen Dualitätssatz gibt es dann weiter ein homogen-dimensionales x mit $\delta(x) = d(y \cdot v)$ und $x \cdot y \cdot v \neq 0$; da $\delta(x) = n - \delta(y) - \delta(v)$, also $\delta(xy) = n - s$ ist, ist $x \cdot y \in \mathfrak{U}_{n-s}$; somit ist v nicht Annullator von $\mathfrak{U}_{n-s}$.

Aus der damit bewiesenen Charakterisierung der minimalen Elemente als

---

d.h. gleich der Dimension der maximalen Abelschen Untergruppen. Die Aufgabe, die Gleichheit zwischen dem Rang einer Gruppen-Mannigfaltigkeit—in dem Sinne, wie wir l oben für jede Mannigfaltigkeit M erklärt haben—und der genannten Dimensionszahl auf möglichst rein geometrischem Wege aufzuklären, habe ich in einer Arbeit behandelt, die in der Com. Math. Helvet. 1941 erscheint.

[31] Alle homogen 1-dimensionalen Elemente sind offenbar minimal.

Annullatoren ist erstens ersichtlich, daß sie eine additive *Gruppe* $\mathfrak{B}_s$ bilden—(dies kann man auch direkt im Anschluß an die Definition in Nr. 32 leicht feststellen)—, und zweitens erkennt man mit Hilfe des Poincaré-Veblenschen Dualitätssatzes jetzt ohne Mühe, daß der Rang dieser Gruppe $\mathfrak{B}_s$ gleich $p_{n-s} - q_{n-s}$ (Nr. 30), also gleich l_s ist. Damit ist der folgende Satz bewiesen, der eine neue Charakterisierung der Zahlen l_s enthält:

Die s-dimensionalen minimalen Elemente von $\mathfrak{R}(M)$ bilden eine additive Gruppe vom Range l_s.

34. *Ein Invarianzsatz.* Die Eigenschaft der "Minimalität" ist invariant gegenüber beliebigen stetigen Abbildungen; genauer: M und M' seien beliebige Mannigfaltigkeiten, und F sei eine Abbildung von M in M'; die dadurch bewirkte Abbildung von $\mathfrak{R}(M)$ in $\mathfrak{R}(M')$ nennen wir ebenfalls F. *Dann ist für jedes minimale Element v von $\mathfrak{R}(M)$ das Bild $F(v)$ minimales Element von $\mathfrak{R}(M')$.*

Beweis: x sei ein homogen-dimensionales Element von $\mathfrak{R}(M)$, und das Element $F(x)$ von $\mathfrak{R}(M')$ sei nicht minimal; wir haben zu zeigen, daß x nicht minimal ist. Falls $d(x) = 0$ ist, ist nichts zu beweisen; es sei $d(x) > 0$; dann ist, da die Abbildung F von $\mathfrak{R}(M)$ in $\mathfrak{R}(M')$ dimensionstreu ist, auch $d(F(x)) > 0$; da $F(x)$ nicht minimal ist, gibt es ein solches homogen-dimensionales Element z von $\mathfrak{R}(M')$, daß $0 < \delta(z) < d(F(x)) = d(x)$ und $z \cdot F(x) \neq 0$ ist. Bezeichnet Φ den Umkehrungs-Homomorphismus von F (Nr. 25), so folgt aus (4.1), daß auch $\Phi(z) \cdot x \neq 0$ ist; dies bedeutet, da nach Nr. 25, 3), $\delta(\Phi(z)) = \delta(z)$, also $0 < \delta(\Phi(z)) < d(x)$ ist: x ist nicht minimal.

35. *Sphärenbilder.* Für die s-dimensionale Sphäre S_s ist dasjenige Element von $\mathfrak{R}(S_s)$, das durch den Grundzyklus repräsentiert wird—also das Eins-Element—minimal; daher ist in dem soeben bewiesenen Satz der folgende enthalten—(statt M' schreiben wir M)—: *Für eine beliebige Mannigfaltigkeit M ist ein Element von $\mathfrak{R}(M)$, das durch das stetige Bild einer Sphäre in M repräsentiert wird, stets ein minimales Element.*[32]

Hieraus folgt weiter, da nach Nr. 33 die Gruppe $\mathfrak{B}_s$ den Rang l_s hat: *Ist $l_s = 0$, so ist in M jedes stetige Bild der s-dimensionalen Sphäre homolog* 0.[32]

36. *Anwendung auf Γ-Mannigfaltigkeiten.* Nach Satz Ib gibt es in einer Γ-Mannigfaltigkeit kein maximales Element z mit geradem $\delta(z)$; daher sind die Ränge l_s für alle geraden s gleich 0; aus Nr. 35 folgt mithin der *Satz III* (Nr. 9).[32]

Für ein Sphärenprodukt

$$S_{m_1} \times S_{m_2} \times \cdots \times S_{m_l}, \qquad d(S_{m_i}) = m_i,$$

gibt die Zahl l_s an, wieviele der Zahlen m_i gleich s sind; dies folgt aus der Charakterisierung der l_s am Schluß von Nr. 31 und der Tatsache, die man leicht

[32] An die Stelle einer wirklichen Sphäre darf offenbar auch eine Homologie-Sphäre (in bezug auf den rationalen Koeffizientenbereich) treten, d.h. eine Mannigfaltigkeit, welche dieselben Bettischen Zahlen hat wie eine Sphäre.

bestätigt, daß die Elemente $Z_1, \cdots, Z_l$, die in Nr. 3, Formel (*), angegeben sind, ein irreduzibles Erzeugenden-System bilden, und daß $\delta(Z_i) = m_i$ ist. Für eine Γ-Mannigfaltigkeit Γ gibt also l_s an, wieviele Faktor-Sphären in dem Sphärenprodukt, das einen mit Γ isomorphen Ring hat, s-dimensional sind. Hierin ist enthalten, daß die Exponenten l_s in Nr. 4, Formel (1′), mit unseren Rängen l_s übereinstimmen.

37. *Die Pontrjaginsche Multiplikation und eine Vermutung.* Die Grundlage für Pontrjagins Untersuchung der Homologie-Eigenschaften geschlossener Liescher Gruppen[15] ist die folgende "Produktbildung": In der Mannigfaltigkeit M sei eine stetige Multiplikation erklärt; x und y seien zwei Zyklen in M; werden x und y von zwei Punkten p bzw. q durchlaufen, so durchläuft das stetige Produkt pq einen Zyklus, den wir mit $x \bigcirc y$ bezeichnen wollen; variieren x und y in ihren Homologieklassen, so ändert sich die Homologieklasse von $x \bigcirc y$ nicht; wir dürfen daher $x \bigcirc y$ als Produkt zweier Homologieklassen x und y deuten; in unserer Ausdrucksweise aus Nr. 26 ist

$$x \bigcirc y = F(x \times y).$$

Diese Produktbildung ist mit der Addition der Homologieklassen offenbar distributiv verknüpft; nehmen wir nun weiter an, die stetige Multiplikation in M sei *assoziativ*, dann wird auch die Produktbildung $x \bigcirc y$ assoziativ sein. Somit sieht man: Für eine *Gruppen*-Mannigfaltigkeit Γ läßt sich die additive Gruppe der Homologieklassen nicht nur durch die Schnittbildung $x \cdot y$ zu dem Ring $\mathfrak{R}(\Gamma)$, sondern außerdem durch die Pontrjaginsche Produktbildung $x \bigcirc y$ zu einem Ring $\mathfrak{P}(\Gamma)$ machen.

Die Bedeutung des Ringes $\mathfrak{P}(\Gamma)$ liegt auf der Hand: in seiner Struktur äußert sich die Wirkung der Gruppen-Multiplikation auf die Zyklen—also ein Zusammenhang zwischen den algebraischen Eigenschaften der Gruppe einerseits, den geometrischen Eigenschaften der Mannigfaltigkeit andererseits. Die Untersuchung dieses Ringes ist daher gewiß eine wichtige Aufgabe. Ich will hierzu eine Vermutung äußern, die sowohl durch Pontrjagins Ergebnisse, als auch durch unseren Satz I nahegelegt wird.[33]

Auch in jedem Sphären-Produkt

$$H = S_{m_1} \times S_{m_2} \times \cdots \times S_{m_l}$$

läßt sich neben der Schnittbildung in naheliegender Weise noch eine zweite Multiplikation, die "Aufspannung", erklären. Wie in Nr. 3 bezeichnen wir den Grundzyklus der Sphäre S_m selbst mit S_m und mit p immer einen einfach gezählten Punkt; wir stellen den Elementen (*) aus Nr. 3 die folgenden Elemente gegenüber:

$$(**) \qquad \begin{aligned} V_1 &= S_{m_1} \times p \times p \times \cdots \times p, \\ V_2 &= p \times S_{m_2} \times p \times \cdots \times p, \\ &\cdots\cdots\cdots\cdots \\ V_l &= p \times p \times p \times \cdots \times S_{m_l}. \end{aligned}$$

[33] Man vergl. zum Folgenden: Cartan[9], (b), 25–26.

Dann definieren wir zunächst für

$$i_1 < i_2 < \cdots < i_k$$

das von $V_{i_1}, V_{i_2}, \cdots, V_{i_k}$ aufgespannte Element

$$V_{i_1} \otimes V_{i_2} \otimes \cdots \otimes V_{i_k} = x_1 \times x_2 \times \cdots \times x_l \tag{5.1}$$

durch die Festsetzung:

$$x_j = S_{m_j} \text{ für } j = i_1, i_2, \cdots, i_k,$$

$$x_j = p \quad \text{für alle anderen } j.$$

Durch (5.1) sind 2^l Elemente erklärt, die eine volle additive Basis bilden. Für je zwei Elemente (5.1)

$$X = V_{i_1} \otimes V_{i_2} \otimes \cdots \otimes V_{i_k}, \qquad Y = V_{j_1} \otimes V_{j_2} \otimes \cdots \otimes V_{j_m} \tag{5.2}$$

definieren wir das aufgespannte Element $X \otimes Y$ dadurch, daß wir die rechten Seiten von (5.2) formal nach dem assoziativen Gesetz miteinander multiplizieren und die Regeln

$$V_j \otimes V_i = (-1)^{m_i m_j} V_i \otimes V_j, \qquad V_i \otimes V_i = 0$$

anwenden; dann wird entweder $X \otimes Y = 0$ oder $\pm X \otimes Y$ wieder ein Element (5.1). Schließlich erklären wir für beliebige Elemente X, Y das Produkt $X \otimes Y$ dadurch, daß wir X und Y als lineare Verbindungen mit rationalen Koeffizienten der Elemente (5.1) darstellen und die distributiven Gesetze anwenden.

Durch die damit erklärte Multiplikation wird die additive Gruppe der Homologieklassen des Sphärenproduktes Π zu einem Ring $\mathfrak{Q}(\Pi)$ gemacht. Die Struktur dieses Ringes ist leicht zu übersehen: man stellt leicht fest, daß die Ringe $\mathfrak{R}(\Pi)$ und $\mathfrak{Q}(\Pi)$ miteinander isomorph sind, und daß dieser Isomorphismus durch eine gewisse Dualität vermittelt wird, die sich zunächst darin äußert, daß einem Element x von $\mathfrak{R}(\Pi)$ immer ein Element y von $\mathfrak{Q}(\Pi)$ mit $d(y) = \delta(x)$ entspricht.

Die oben erwähnte *Vermutung* bezüglich des Pontrjaginschen Ringes $\mathfrak{P}(\Gamma)$ einer *Gruppen*-Mannigfaltigkeit Γ ist nun die folgende:

Die, auf Grund des Satzes I mögliche, isomorphe Abbildung der Ringe $\mathfrak{R}(\Gamma)$ und $\mathfrak{R}(\Pi)$ aufeinander läßt sich so wählen, daß sie zugleich die Ringe $\mathfrak{P}(\Gamma)$ und $\mathfrak{Q}(\Pi)$ isomorph aufeinander abbildet.

Damit wäre die Struktur des Ringes $\mathfrak{P}(\Gamma)$ sowie die Beziehung zwischen den Ringen $\mathfrak{P}(\Gamma)$ und $\mathfrak{R}(\Gamma)$ weitgehend geklärt.

Unter anderem würde noch die folgende Rolle der minimalen Elemente einer Gruppen-Mannigfaltigkeit sichtbar werden. Offenbar sind die Elemente $V_1, \cdots, V_l$, die durch (**) gegeben sind, minimale Elemente von $\mathfrak{R}(\Pi)$, und zwar bilden sie eine additive Basis der *vollen* Gruppe der minimalen Elemente von $\mathfrak{R}(\Pi)$, d.h. der Vereinigung der Gruppen $\mathfrak{B}_1, \cdots, \mathfrak{B}_n$ (Nr. 33); daher würden auch die Elemente $v_1, \cdots, v_l$ von $\mathfrak{R}(\Gamma)$, welche bei dem vermuteten Isomorphismus den V_i entsprächen, minimal sein und eine Basis der vollen

Gruppe der minimalen Elemente von $\mathfrak{R}(\Gamma)$ bilden; da aber die Elemente (5.1) eine volle additive Basis in Π bilden, würden auch die Pontrjaginschen Produkte

$$v_{i_1} \bigcirc v_{i_2} \bigcirc \cdots \bigcirc v_{i_k},$$
$$i_1 < i_2 \cdots < i_k,$$

eine volle additive Basis in Γ bilden. Es würde sich also herausstellen, daß die minimalen Elemente $v_1, v_2, \cdots, v_l$ in ganz analoger Weise durch die Pontrjaginsche Produktbildung den Ring $\mathfrak{P}(\Gamma)$ aufspannen, wie die maximalen Elemente $z_1, z_2, \cdots, z_l$ durch die Schnittbildung den Ring $\mathfrak{R}(\Gamma)$ erzeugen.[34]

ZÜRICH, SWITZERLAND.

[34] Die oben ausgesprochene Vermutung, für die ich keinen Beweis gefunden hatte, habe ich mündlich Herrn H. Samelson mitgeteilt, und dieser hat ihre Richtigkeit inzwischen in vollem Umfange bewiesen.—Sie bezieht sich übrigens ausdrücklich auf Gruppenmannigfaltigkeiten, und für den Beweis ist die Benutzung des assoziativen Gesetz wesentlich.

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

ON THE CARTAN INVARIANTS OF GROUPS OF FINITE ORDER*

BY RICHARD BRAUER

(Received April 18, 1940)

1. INTRODUCTION

E. Cartan, in his fundamental paper on hypercomplex numbers,[1] introduced an important set of invariants $c_{\kappa\lambda}$, $(\kappa, \lambda = 1, 2, \cdots, k)$ of an algebra A with a principal unit 1. Here, k is the number of prime ideals $\mathfrak{P}_\kappa$ of A. The $c_{\kappa\lambda}$ are non-negative integers which also play an important rôle in the decomposition of the regular representation of A.

Let us consider now a semisimple algebra Γ of rank n over an algebraic number field K, and let J be an integral domain[2] of Γ. Every prime ideal $\mathfrak{p}$ of K generates an ideal $\mathfrak{p}_J$ of J. The nature of this ideal is determined by the structure of the residue class ring $J/\mathfrak{p}_J$. This ring can be considered as an algebra over the residue class ring $\mathfrak{o}/\mathfrak{p}$, where $\mathfrak{o}$ denotes the domain of all integers of K. Hence, we may form the Cartan invariants $c_{\kappa\lambda}(\mathfrak{p})$ of $J/\mathfrak{p}_J$. We have in this case $c_{\kappa\lambda}(\mathfrak{p}) = c_{\lambda\kappa}(\mathfrak{p})$, and the $c_{\kappa\lambda}(\mathfrak{p})$ are the coefficients of a non-negative quadratic form $\psi = \sum c_{\kappa\lambda}(\mathfrak{p})x_\kappa x_\lambda$.

In particular, these notions can be used in the case of the group ring Γ of a group $\mathfrak{G}$ of finite order g. As field of reference, we choose an algebraic number field K, such that all the absolutely irreducible representations of $\mathfrak{G}$ can be written with coefficients in K.[3] The linear combinations of the group elements with integral coefficients in K form an integral domain J of Γ. Hence, we may form the Cartan invariants $c_{\kappa\lambda}(\mathfrak{p})$ for every prime ideal $\mathfrak{p}$ of K. It appears that they actually depend only on the rational prime p which is divisible by $\mathfrak{p}$. Accordingly, we denote them by $c_{\kappa\lambda}(p)$.[4] If p is not a divisor of the group order g, then the matrix $C(p) = (c_{\kappa\lambda}(p))$ is the unit matrix 1, i.e. $c_{\kappa\lambda}(p) = \delta_{\kappa\lambda}$. We therefore restrict our attention to the case where p divides g, in which case p is a divisor of the discriminant of J. To every such prime p, we obtain in $c_{\kappa\lambda}(p)$

* Presented to the American Mathematical Society on April 26, 1940.

[1] E. Cartan, Annales de Toulouse, **12** B, (1898), p. 1. Cf. further R. Brauer, Proc. Nat. Acad. Sci. **25**, (1939), p. 252.

[2] By an integral domain (Ordnung) J of Γ, we understand a subring J of Γ with the following properties: (a) J contains all the integers of K; (b) The rank of J is n; (c) The elements of J, when expressed by a basis $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ of J, have the form $\alpha = \sum \frac{a_i}{w} \epsilon_i$, where the a_i are integers of K and w is a fixed integer of K which is independent of α.

[3] Cf., for instance, A. Speiser, Theorie der Gruppen von endlicher Ordnung, 3 rd ed., Berlin 1937, theorem 181, p. 204.

[4] For the properties of the $c_{\kappa\lambda}(p)$, in this case, cf. R. Brauer and C. Nesbitt, University of Toronto Studies, Math. Series No. 4, 1937, and a paper forthcoming in the Ann. of Math.

a set of invariants of the group $\mathfrak{G}$, which are of great importance for the theory of group characters. The aim of this paper is the determination of the discriminant $|C(p)|$ of ψ. We prove

THEOREM 1: *The determinant* $|c_{\kappa\lambda}(p)|$ *of the matrix of Cartan invariants of a group* $\mathfrak{G}$ *of finite order is a power of* p.

The exact exponent of p in $|c_{\kappa\lambda}(p)|$ is given below in theorem 1*.

2. Preparations for the Proof

We denote by $\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_k$, the classes of conjugate elements of the group $\mathfrak{G}$, which consist of p-regular elements.[5] Let $g = g/n_\lambda$ be the number of elements in $\mathfrak{C}_\lambda$, so that n_λ is the order of the normalizor of an element of $\mathfrak{C}_\lambda$. To each class $\mathfrak{C}_\lambda$, there corresponds a reciprocal class $\mathfrak{C}_{\lambda^*}$ containing the reciprocals of the elements of $\mathfrak{C}_\lambda$. We have, then, k absolutely irreducible modular characters, $\varphi^{(1)}, \varphi^{(2)}, \cdots, \varphi^{(k)}$ of $\mathfrak{G}$ (mod p); we may arrange the values $\varphi_\lambda^{(\kappa)}$ of $\varphi^{(\kappa)}$ for the class $\mathfrak{C}_\lambda$ in the form of a matrix

$$\Phi = (\varphi_\lambda^{(\kappa)}) \qquad (\kappa, \lambda = 1, 2, \cdots, k).$$

The number k here also gives the degree of the Cartan matrix $C = (c_{\kappa\lambda}(p))$ and the matrices Φ and C are connected by the formula

$$\Phi' C \Phi = (n_\kappa \delta_{\kappa\lambda^*}),^{6} \tag{1}$$

provided that $\varphi^{(1)}, \varphi^{(2)}, \cdots, \varphi^{(k)}$ are taken in a suitable arrangement. On forming the determinant in (1), we obtain

$$|\Phi|^2 |C| = \pm n_1 n_2 \cdots n_k . \tag{2}$$

This shows that $|C| \neq 0$. Hence, the non-negative quadratic form ψ is positive definite, i.e. $|C| > 0$. Further, the determinant $|\Phi|$ is prime to p.[7] Therefore, theorem 1 will be proved when we can show

THEOREM 2: *Let* $\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_k$ *be the classes of p-regular, conjugate elements of* $\mathfrak{G}$, *and let* $g_\lambda = g/n_\lambda$ *be the number of elements of* $\mathfrak{C}_\lambda$. *If* Φ *is the matrix of the modular group characters of* $\mathfrak{G}$(mod p), *then the square of the determinant* $|\Phi|$ *is given by*

$$|\Phi|^2 = \pm \frac{n_1 n_2 \cdots n_k}{p^\alpha} \tag{3}$$

where p^α *is the highest power of* p *dividing* $n_1 n_2 \cdots n_k$.

At the same time, we obtain

THEOREM 1*: *The determinant of the Cartan matrix,* $(c_{\kappa\lambda}(p))$, *is equal to* p^α, *where* p^α *has the same significance as in theorem* 2.

For primes p which do not divide the order of $\mathfrak{G}$, theorem 2 is trivial. Here,

[5] By a p-regular element of $\mathfrak{G}$, we understand an element whose order is prime to p.

[6] Cf. the formulae (29) and (15), respectively, of the papers mentioned in [4].

[7] Cf. the formulae (26), (17), respectively, of the papers mentioned in [4].

k is the number of all the classes of conjugate elements of $\mathfrak{G}$, and $p^\alpha = 1$. The relation (3) is obtained by multiplying Φ' and Φ, using the orthogonality relations for group characters. In the same manner, the analogous formula for ordinary group characters (instead of modular characters) can be obtained at once.

In order to prove theorem 2, it is sufficient to show that if $q \neq p$ is a rational prime, and if q^β divides the right hand side of (2), then q^β divides $|\Phi|^2$. We shall prove that by proving a similar statement for certain minors of Φ. We first have to give some simple group theoretical considerations.

Let A be an element of $\mathfrak{G}$ such that the order of A is prime to p and q. We shall say that an element G of $\mathfrak{G}$ contains A as its *q-regular factor*, if G is of the form $G = AQ$, where the order of Q is a power $q^\nu \geqq 1$ of q, and where $AQ = QA$. Of course, A and Q are uniquely determined by G; both can be written as powers of G. If G_1 is conjugate to AQ, then the q-regular factor of G_1 is conjugate to A.

Let $A_1, A_2, \cdots, A_m$ be a maximal system of elements of $\mathfrak{G}$, such that A_i, A_j are not conjugate for $i \neq j$ and the order of each A_i is prime to p and q. With each A_i, we associate those classes of conjugate elements, $\mathfrak{C}_1^{(i)}, \mathfrak{C}_2^{(i)}, \cdots, \mathfrak{C}_{h_i}^{(i)}$, which contain elements with A_i as their q-regular factor. Each of the classes, $\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_k$, then appears exactly once in the form $\mathfrak{C}_\mu^{(i)}$. By expanding $|\Phi|$, we see that $|\Phi|$ is a sum of terms

$$T_1 T_2 \cdots T_m, \tag{4}$$

where T_i is a minor of degree h_i of Φ, containing only the columns which belong to $\mathfrak{C}_1^{(i)}, \mathfrak{C}_2^{(i)}, \cdots, \mathfrak{C}_{h_i}^{(i)}$.

We now state

THEOREM 3: *Let A be an element of an order not divisible by the two primes p and q, and assume that the h classes, $\mathfrak{C}_\rho, \mathfrak{C}_\sigma, \cdots, \mathfrak{C}_\tau$, are all the classes of conjugate elements of the group $\mathfrak{G}$, which contain elements G with A as their q-conjugate factor. If $\varphi^{(r)}, \varphi^{(s)}, \cdots, \varphi^{(t)}$ are any h modular characters* (mod p) *of $\mathfrak{G}$, then*

$$|\Delta| = \begin{vmatrix} \varphi_\rho^{(r)} & \varphi_\sigma^{(r)} & \cdots & \varphi_\tau^{(r)} \\ \varphi_\rho^{(s)} & \varphi_\sigma^{(s)} & \cdots & \varphi_\tau^{(s)} \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_\rho^{(t)} & \varphi_\sigma^{(t)} & \cdots & \varphi_\tau^{(t)} \end{vmatrix} \equiv 0 \pmod{(q_\rho q_\sigma \cdots q_\tau)^{\frac{1}{2}}}, \tag{5}$$

where q_λ is the highest power of q dividing n_λ.[8]

Each T_i in (4) has the form (5) (for $A = A_i$). From theorem 3, it follows that the expression (4) is divisible by $(q_1 q_2 \cdots q_k)^{\frac{1}{2}}$, and, hence, that $|\Phi|$ is divisible by the same number. It is therefore sufficient to prove theorem 3 in order to prove theorems 1 and 2. Changing the notation, if necessary, we may assume without restriction that

$$\rho = 1, \sigma = 2, \cdots, \tau = h,$$
$$r = 1, s = 2, \cdots, t = h,$$

[8] An analogous theorem holds for ordinary group characters. Here, the assumption that the order of A is prime to p is not necessary. The proof is the same as for theorem 3.

and then (5) assumes the form

$$(6) \qquad |\Delta| = \begin{vmatrix} \varphi_1^{(1)}, & \cdots, & \varphi_h^{(1)} \\ \cdots & \cdots & \cdots \\ \varphi_1^{(h)}, & \cdots, & \varphi_h^{(h)} \end{vmatrix} \equiv 0 \pmod{(q_1 q_2 \cdots q_h)^{\frac{1}{2}}}.$$

The proof of (6) will be given in §3. First, we must formulate and prove a group theoretical lemma. Let $AQ_1, AQ_2, \cdots, AQ_h$ be representatives for the h classes $\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_h$, where A is the q-regular factor of AQ_i, and where $Q_1 = 1$. Let $\mathfrak{N}$ be the normalizor of A in $\mathfrak{G}$, and let $\mathfrak{Q}$ be a Sylow subgroup of $\mathfrak{N}$ belonging to the prime q. Then n_1 is the order of $\mathfrak{N}$ and q_1 the order of $\mathfrak{Q}$. Each AQ_i will commute with every element of a certain subgroup $\mathfrak{Q}_i$ of order q_i of $\mathfrak{G}$. Since A and Q_i both are powers of AQ_i, each of them commutes with every element of $\mathfrak{Q}_i$.

In particular, we have $\mathfrak{Q}_i \subseteq \mathfrak{N}$. Replacing Q_i by an element $N_i^{-1}Q_iN_i$, with N_i in $\mathfrak{N}$, we may assume that

$$(7) \qquad \mathfrak{Q}_i \subseteq \mathfrak{Q},$$

as follows easily from Sylow's theorem. Since Q_i must belong to $\mathfrak{Q}_i$, the element Q_i itself will belong to $\mathfrak{Q}$. If Q is any element of $\mathfrak{Q}$, then AQ will be conjugate in $\mathfrak{G}$ to some AQ_i, i.e., $G^{-1}AQG = AQ_i$. Raising this equation to suitable exponents, we obtain $G^{-1}AG = A$, $G^{-1}QG = Q_i$. Therefore, Q and Q_i are conjugate in $\mathfrak{N}$, and hence $Q_1, Q_2, \cdots, Q_h$ form a complete system of representatives for those classes of conjugate elements in $\mathfrak{N}$, in which the orders of the elements are powers of q.

In $\mathfrak{Q}$, the elements $Q_1, Q_2, \cdots, Q_h$ need not form a complete system of representatives for the classes of conjugate elements. However, we may construct such a system by adding further elements Q to the set $Q_1, Q_2, \cdots, Q_h$. Each Q will, in $\mathfrak{N}$, be conjugate to a certain Q_i where i is uniquely determined, $i = 1, 2, \cdots, h$. We denote the elements Q belonging to Q_i by $Q_i = Q_i^{(0)}, Q_i^{(1)}, Q_i^{(2)}, \cdots, Q_i^{(l_i)}$, $(l_i \geqq 0)$. Let $q_i^{(\lambda)}$ be the highest power of q dividing the order of the normalizor of $Q_i^{(\lambda)}$ in $\mathfrak{Q}$. According to (7), we have

$$(8) \qquad q_i^{(0)} = q_i.$$

We now prove the

LEMMA: *The numbers $q_i^{(\lambda)}$, $(\lambda = 0, 1, 2, \cdots, l_i)$ are divisors of q_i. If exactly d_i of them are equal to q_i, then*

$$(9) \qquad d_i \not\equiv 0 \pmod{q}.$$

PROOF: Let $\mathfrak{K}_i$ denote the class of conjugate elements of $\mathfrak{N}$, which contains the element AQ_i. The number M_i of elements in $\mathfrak{K}_i$ is equal to the order of $\mathfrak{N}$ divided by the order of the normalizor of AQ_i. Hence

$$(10) \qquad M_i = \frac{q_1}{q_i} \tilde{M}_i, \qquad \text{with } (\tilde{M}_i, q) = 1.$$

The class $\mathfrak{K}_i$ can be broken up into partial classes $\mathfrak{K}_i^{(\mu)}$, where each $\mathfrak{K}_i^{(\mu)}$ consists of elements which are conjugate by means of transformations by elements of $\mathfrak{Q} \subseteq \mathfrak{N}$. The elements $AQ_i^{(0)}, AQ_i^{(1)}, \cdots, AQ_i^{(l_i)}$ will each determine such a partial class, but there may be further partial classes which do not contain elements AQ with Q in $\mathfrak{Q}$. In any case, if $AT_i^{(\mu)}$ is an element of $\mathfrak{K}_i^{(\mu)}$, and if $T_i^{(\mu)}$ commutes with exactly $w_i^{(\mu)}$ elements of $\mathfrak{Q}$, then the number $M_i^{(\mu)}$ of elements of $\mathfrak{K}_i^{(\mu)}$ is given by

$$M_i^{(\mu)} = \frac{q_1}{w_i^{(\mu)}}. \tag{11}$$

We have, of course,

$$M_i = \sum_{\mu} M_i^{(\mu)}. \tag{12}$$

In $\mathfrak{G}$, the elements AQ_i and $AT_i^{(\mu)}$ are conjugate. Hence, the order of the normalizor of $AT_i^{(\mu)}$ in $\mathfrak{G}$ is divisible by q_i but not by a higher power of q. On considering the subgroup generated by $AT_i^{(\mu)}$ and the $w_i^{(\mu)}$ commuting elements of $\mathfrak{Q}$, we readily see that $w_i^{(\mu)} \leqq q_i$ and that the equality sign can hold only if $T_i^{(\mu)}$ belongs to $\mathfrak{Q}$. When $T_i^{(\mu)} = Q_i^{(\lambda)}$, $w_i^{(\mu)} = q_i^{(\lambda)}$, and we thus obtain the first part of the lemma. If $w_i^{(\mu)} = q_i$, then the partial class $\mathfrak{K}_i^{(\mu)}$ contains exactly one element $Q_i^{(\lambda)}$, $(\lambda = 0, 1, 2, \cdots, l_i)$, and we have $q_i^{(\lambda)} = q_i$. According to our assumption, there are exactly d_i such partial classes. Therefore, d_i of the numbers $M_i^{(\mu)}$, (cf. (11)), are equal to q_1/q_i, the remaining ones being divisible by a higher power power of q. Then (12) gives

$$M_i \equiv d_i \frac{q_1}{q_i} \qquad \left(\operatorname{mod} \frac{qq_1}{q_i}\right)$$

and, on comparing this with (10), we obtain (9).

3. Proof of the Theorems

As we have seen, it is sufficient to prove (6). If $\mathfrak{F}$ is any representation of $\mathfrak{G}$, we may assume that the matrix $\mathfrak{F}_A$ representing the fixed element A appears in canonical form, i.e.

$$A \rightarrow \mathfrak{F}_A = \begin{pmatrix} \alpha_1 I_{v_1} & & \\ & \alpha_2 I_{v_2} & \\ & & \ddots \end{pmatrix}$$

where $\alpha_1, \alpha_2, \cdots$ are distinct roots of unity and $v_1, v_2, \cdots$ are positive integers. The matrices $\mathfrak{F}_Q$, representing elements Q of $\mathfrak{Q}$, then break up in the form

$$Q \rightarrow \mathfrak{F}_Q = \begin{pmatrix} V_1 & & \\ & V_2 & \\ & & \ddots \end{pmatrix}$$

where V_j is of degree v_j. For a fixed j, the matrices V_j form a representation $\mathfrak{B}_j$ of $\mathfrak{Q}$. Let $\vartheta^{(1)}, \vartheta^{(2)}, \cdots, \vartheta^{(m)}$ be the distinct, irreducible characters[9] of $\mathfrak{Q}$. Then m is the number of classes of conjugate elements in $\mathfrak{Q}$, i.e., the number of elements $Q_i^{(\lambda)}$. If χ denotes the character of $\mathfrak{F}$, we readily obtain

$$\chi(AQ) = \sum_{\mu=1}^{m} z_\mu \vartheta^{(\mu)}(Q)$$

where the z_μ are algebraic integers which are independent of Q. We set, accordingly,

$$\varphi^{(\kappa)}(AQ) = \sum_\mu z_{\kappa\mu}\vartheta^{(\mu)}(Q). \tag{13}$$

In particular, we use this for $Q = Q_i^{(\nu)}$. Since Q_i and $Q_i^{(\nu)}$ are conjugate in $\mathfrak{N}$, we have

$$\sum_\mu z_{\kappa\mu}\vartheta^{(\mu)}(Q_i) = \sum_\mu z_{\kappa\mu}\vartheta^{(\mu)}(Q_i^{(\nu)}). \tag{14}$$

We now introduce matrix notation. Let us denote by Θ the matrix $(\vartheta^{(\kappa)}(Q_i^{(\nu)}))$ of the characters of $\mathfrak{Q}$; the rows here are fixed by one index κ, $(\kappa = 1, 2, \cdots, m)$, the columns by the two indices i, ν, $(i = 1, 2, \cdots, h;\ \nu = 0, 1, 2, \cdots, l_i)$. We arrange the columns so that first the h columns with $\nu = 0$ appear and then the $m - h$ other columns. Thus,

$$\Theta = (\Theta_0, \Theta_1)$$

where Θ_0 is of type (m, h)[10] and Θ_1 of type $(m, m - h)$.

Since the notation in (6) was chosen such that the class $\mathfrak{C}_\lambda$ contained AQ_λ, we have $\varphi_\lambda^{(\kappa)} = \varphi^{(\kappa)}(AQ_\lambda)$. On account of (13), the matrix Δ in (6) can be written in the form

$$\Delta = Z\Theta_0$$

where Z is the matrix $(z_{\kappa\lambda})$, $(\kappa = 1, 2, \cdots, h;\ \lambda = 1, 2, \cdots, m)$, of type (h, m). In order to replace the matrices by square matrices, we set

$$\tilde{Z} = \begin{pmatrix} Z \\ U \end{pmatrix}$$

where U is a matrix of type $(m - h, m)$ with arbitrary integral coefficients. Then

$$\tilde{Z}\Theta = \begin{pmatrix} Z \\ U \end{pmatrix}(\Theta_0, \Theta_1) = \begin{pmatrix} Z\Theta_0 & Z\Theta_1 \\ U\Theta_0 & U\Theta_1 \end{pmatrix} = \begin{pmatrix} \Delta & Z\Theta_1 \\ U\Theta_0 & U\Theta_1 \end{pmatrix}.$$

Here we subtract every column $(i, 0)$ from all the columns (i, λ), with $\lambda > 0$, and with the same first index. Then, according to (14), $Z\Theta_1$ on the right hand

[9] Since the order of $\mathfrak{Q}$ is prime to p, there is no difference here between the ordinary and the modular characters (mod p) of $\mathfrak{Q}$.

[10] i.e., a matrix with m rows and h columns.

side will be replaced by 0. If Θ_1^* is obtained from Θ_1 by subtracting from each column (i, λ), with $\lambda > 0$, the column $(i, 0)$ of Θ_0, then, by taking the determinant, we find

$$|\tilde{Z}||\Theta| = |\Delta||U\Theta_1^*|.$$

Here, $|\tilde{Z}|$ is an algebraic integer, and $|\Theta|$ has the value

$$|\Theta| = \prod_{i=1}^{h}\prod_{\nu=0}^{l_i}(q_i^{(\nu)})^{\frac{1}{2}}.\ [11]$$

On account of (8), we obtain

$$|\Delta||U\Theta_1^*| \equiv 0 \pmod{(q_1 q_2 \cdots q_k)^{\frac{1}{2}}(\textstyle\prod_i \prod_{\lambda>0} q_i^{(\lambda)})^{\frac{1}{2}}}. \tag{15}$$

Let us assume now that the formula (6) does not hold. Since all the $q_i^{(\mu)}$ are powers of q, it follows that

$$|U\Theta_1^*| \equiv 0 \pmod{(q \textstyle\prod_i \prod_{\lambda>0} q_i^{(\lambda)})^{\frac{1}{2}}}, \tag{16}$$

for any choice of U. Taking a suitable U, we see that any minor of degree $m - h$ of Θ_1^* can be obtained in the form $|U\Theta_1^*|$. But the determinant $|(\Theta_1^*)'\Theta_1^*|$ is equal to the sum of the squares of all these minors. Hence,

$$|(\Theta_1^*)'\Theta_1^*| \equiv 0 \pmod{q \textstyle\prod_i \prod_{\lambda>0} q_i^{(\lambda)}}. \tag{17}$$

Any row in $(\Theta_1^*)'$, the transpose of Θ_1^*, is characterized by a pair of indices, i, μ, $(i = 1, 2, \cdots, h;\ \mu = 1, 2, \cdots, l_i)$, and any column is characterized by an index κ, $(\kappa = 1, 2, \cdots, m)$. The rows of $(\Theta_1^*)'\Theta_1^*$ are given in the same manner, and each column is characterized by a pair of indices j, ν, with $j = 1, 2, \cdots, h$; $\nu = 1, 2, \cdots, l_j$. For the element $y(i, \mu; j, \nu)$ at the place $(i, \mu), (j, \nu)$ in $(\Theta_1^*)'\Theta_1^*$, we obtain easily

$$y(i, \mu; j, \nu) = \sum_{\kappa=1}^{m}(\vartheta^{(\kappa)}(Q_i^{(\mu)}) - \vartheta^{(\kappa)}(Q_i))(\vartheta^{(\kappa)}(Q_j^{(\nu)}) - \vartheta^{(\kappa)}(Q_j))$$

on account of the definition of Θ_1^*. The sum here splits into four sums, each of which can be computed by means of the orthogonality relations for the group characters of $\mathfrak{Q}$. We set $\delta(R, S) = 1$, if the elements R and S^{-1} of $\mathfrak{Q}$ are conjugate in $\mathfrak{Q}$, and in the other case we set $\delta(R, S) = 0$. Since the normalizor of $Q_i^{(\mu)}$ in $\mathfrak{Q}$ has the order $q_i^{(\mu)}$, we find

$$\begin{aligned} &y(i, \mu; j, \nu) \\ &\quad = \delta(Q_i^{(\mu)}, Q_j^{(\nu)})q_i^{(\mu)} - \delta(Q_i, Q_j^{(\nu)})q_i - \delta(Q_i^{(\mu)}, Q_j)q_i^{(\mu)} + \delta(Q_i, Q_j)q_i\end{aligned} \tag{18}$$

For each i, this expression vanishes unless Q_j^{-1} and Q_i are conjugate in $\mathfrak{G}$, since otherwise $Q_i^{(\rho)}$ and $Q_j^{(\sigma)-1}$ could not be conjugate in $\mathfrak{Q}$. Consequently, there is only one value $j = i^*$ for a given i for which the expression can be different

[11] This is the analogue of theorem 2 for ordinary group characters, and, as remarked in connection with theorem 2, this analogue is trivial.

from 0, and we have $l_i = l_{i^*}$. This shows that the determinant (17) splits into a product of m determinants

$$(19) \qquad |(\Theta_1^*)'\Theta_1^*| = \pm|\Omega_1|\cdot|\Omega_2|\cdot\cdots\cdot|\Omega_h|,$$

where

$$(20) \qquad |\Omega_i| = |y(i, \mu; i^*, \nu)|^{12} \quad \begin{cases} i, i^* \text{ fixed, } \mu \text{ row index,} \\ \nu \text{ column index, } \mu, \nu = 1, 2, \cdots, l_i . \end{cases}$$

According to the lemma in §2, each q_i is divisible by $q_i^{(\mu)}$; we set, accordingly,

$$(21) \qquad y(i, \mu, i^*, \nu) = q_i^{(\mu)} x_i(\mu, \nu).$$

Then

$$(22) \qquad |\Omega_i| = q_i^{(1)} q_i^{(2)} \cdots q_i^{(l_i)} |X_i|,$$

where

$$(23) \qquad X_i = (x_i(\mu, \nu)), \qquad (\mu \text{ row-index, } \nu \text{ column-index}).$$

Using (17), (19), and (22), we obtain

$$(24) \qquad |X_1||X_2|\cdots|X_m| \equiv 0 \pmod{q}.$$

Because of (18) and (21), the matrix X_i is the sum of four matrices $X_i^{(1)}, X_i^{(2)}, X_i^{(3)}, X_i^{(4)}$. In each case we have i fixed, and we denote the row-index by μ, the column index by ν.

$$(25) \qquad \begin{cases} X_i^{(1)} = (\delta(Q_i^{(\mu)}, Q_{i^*}^{(\nu)}), & X_i^{(2)} = -\left(\delta(Q_i, Q_{i^*}^{(\nu)}) \dfrac{q_i^{(\mu)}}{q_i}\right), \\ X_i^{(3)} = -(\delta(Q_i^{(\mu)}, Q_{i^*})), & X_i^{(4)} = \left(\delta(Q_i, Q_{i^*}) \dfrac{q_i^{(\mu)}}{q_i}\right). \end{cases}$$

As the lemma in §2 shows, we have $d_i - 1$ values for $\mu \geqq 1$, for which $q_i^{(\mu)} = q_i$. We may assume that these are the values $\mu = 1, 2, \cdots, d_i - 1$.[13] For $\mu \geqq d_i$, we have

$$(26) \qquad q_i/q_i^{(\mu)} \equiv 0 \pmod{q}.$$

Three cases must be considered separately.

Case I: $i \neq i^*$.

We may then assume that $Q_i^{(\mu)-1} = Q_{i^*}^{(\mu)}$ for all μ. Then $X_i^{(1)}$ is the unit matrix, $X_i^{(2)} = X_i^{(3)} = 0$, whereas in $X_i^{(4)}$ each coefficient in the μ-th row is equal to $q_i/q_i^{(\mu)}$. Hence, (mod q), the matrix $X_i^{(4)}$ has $d_i - 1$ rows consisting of 1, and the other rows are all 0. Then

$$X_i^{(4)\prime} X_i^{(4)} \equiv (d_i - 1)X_i^{(4)}, \qquad \operatorname{tr}(X_i^{(4)}) \equiv d_i - 1 \pmod{q}.$$

[12] If $l_i = 0$, we must set $|\Omega_i| = 1$.

[13] For $d_i = 1$, the corresponding kinds of rows of $X_i^{(j)}$, do not occur.

This shows that (mod q) one characteristic root of X has the value $d_i - 1$, and that the others have the value 0. Then the characteristic roots of $X_i = I + X_i^{(4)}$ are given by d_i, 1, 1, $\cdots$, 1 (mod q). Hence

$$(27_1) \qquad | X_i | \equiv d_i \pmod{q}.$$

CASE II. $i = i^*$, but Q_i and Q_i^{-1} are not conjugate in $\mathfrak{Q}$.

We may assume here that $Q_i^{-1} = Q_i^{(1)}$. Then $X_i^{(4)} = 0$. In $X_i^{(2)}$, only the first column contains elements different from 0, and the coefficients in this column are given by

$$-q_i/q_i^{(1)}, \ -q_i/q_i^{(2)}, \ \cdots, \ -q_i/q_i^{(l_i)}.$$

In $X_1^{(3)}$, only the first row contains elements different from 0. All the coefficients in the first row are equal to -1.

In $X_i^{(1)}$, the first row and column are 0. Each of the other rows contains exactly one coefficient 1, and all the other coefficients are 0. The same is true for the second, third, $\cdots$, last column. On adding all the other rows to the first row in X_i, we obtain easily

$$| X_i | = \pm\left(\frac{q_i}{q_i^{(1)}} + \frac{q_i}{q_i^{(2)}} + \cdots + \frac{q_i}{q_i^{(l_i)}} + 1\right).$$

Hence, since $d_i - 1$ of the fractions are 1, and the other ones $\equiv 0 \pmod{q}$,

$$(27_2) \qquad | X_i | \equiv \pm d_i \pmod{q}.$$

CASE III. $i = i^*$, and Q_i and Q_i^{-1} are conjugate in $\mathfrak{Q}$. Here $X_i^{(2)} = 0$, $X_i^{(3)} = 0$. As in case I, the first $d_i - 1$ rows in $X_i^{(4)}$ contain only coefficients $\equiv 1 \pmod{q}$, and the latter rows contain only coefficients $\equiv 0 \pmod{q}$. The matrix $X_i^{(1)}$ can be changed into the unit matrix, if the columns are taken in another order; the value of $X_i^{(4)}$ (mod q) is not altered hereby. The argument used in the first case then gives

$$(27_3) \qquad \pm \, | X_i | \equiv d_i \pmod{q}.$$

The three formulae (27) show, in connection with the lemma in §2, that in any case $| X_i | \not\equiv 0 \pmod{q}$. Then (24) is impossible. Thus, the assumption that (6) is not true leads to a contradiction, and the theorems 1, 2 and 3 are proved.

UNIVERSITY OF TORONTO

Annals of Mathematics
Vol. 42, No. 1, January, 1941

MINIMAL SURFACES NOT OF MINIMUM TYPE BY A NEW MODE OF APPROXIMATION

Marston Morse and C. Tompkins

(Received May 31, 1940)

1. Introduction. We begin with a study of a non-negative function $f^0(p)$ defined at each point of a metric space M. We suppose that $f^0(p)$ is the limit of a sequence of functions $f^n(p)$ also defined over M. Let p_n be a homotopic critical point p_n of $f^n(p)$. We study the limit points q of sequences p_n, regarding q as a new type of critical point of the limit function $f^0(p)$.

We give a simple application.

Let g^0 be a simple, closed, rectifiable curve in a euclidean space with the property that the ratio of an arbitrary chord length of g^0 to the smaller of the corresponding two arc lengths of g^0 is bounded from zero. This chord arc condition on g^0 is less restrictive than the condition imposed by Shiffman[1] in his study of unstable minimal surfaces. We make use of Douglas's Dirichlet function $f^0(\varphi)$. The point φ is a monotone transformation of the arc length along g^0. The basic result is as follows.

If g^0 bounds two minimal surfaces of disc type belonging to disjoint minimizing sets of such surfaces (cf. §2), *then g^0 also bounds another minimal surface of disc type not of minimum type.*

Results of this nature but under more restrictive hypotheses have been obtained previously by Morse and Tompkins[2] and by Shiffman, loc. cit. We have found that less restrictive hypotheses on g^0 are adequate, because they can be compensated by more restrictive hypotheses on the approximating functions $f^n(p)$, $n > 0$, making the analysis of the functions $f^n(p)$ much simpler. This applies particularly to the study of upper-reducibility.[3] For it is easy to show that the restricted functions $f^n(p)$ are upper-reducible, while we do not need to show that $f^0(p)$ is upper-reducible.

2. The mode of approximation. Let M be a metric space with points p, q, r and distances pq, pr, etc. satisfying the usual metric axioms. On M there shall be given a sequence of functions $f^n(p)$ converging to a function $f^0(p)$. Thus

$$\lim f^n(p) = f^0(p). \tag{2.1}$$

[1] Shiffman. *The Plateau problem for non-relative minima.* Annals of Mathematics, vol. 40 (1939), pp. 834–854. See (2) §3.

[2] Morse and Tompkins. *The existence of minimal surfaces of general critical types.* Annals of Mathematics, vol. 40 (1939), pp. 443–472. The letters MT will be used in the text to refer to this paper. See corrections elsewhere in this number of the Annals.

[3] Morse. *Functional topology and abstract variational theory.* Mémorial des Sciences Mathématiques, Fascicule 92 (1939). See §8. The letter M will refer to this pamphlet.

This convergence is in general not uniform. The functions $f^m(p)$, $m = 0, 1, \cdots$, shall satisfy conditions I to VI to be described. If c is an arbitrary constant and f is a function defined on M, f_c shall denote the subset of points of M at which $f \leqq c$. Our first condition is as follows.

HYPOTHESIS I. *Bounded compactness. The sets* f_c^m, $m = 0, 1, \cdots$, *shall be compact for each constant* c.

It follows from this hypothesis that each function f^m is lower semi-continuous.

HYPOTHESIS II. *If* $\lim p_n = p$,

$$\text{inf. lim. } f^n(p_n) \geqq f^0(p). \tag{2.2}$$

HYPOTHESIS III. *Corresponding to any subsequence F^r of the functions f^n and points p_r such that $F^r(p_r)$ is bounded independently of r, there exists a subsequence of the points p_r which converges to a point of M.*

We shall prove the following lemma.

LEMMA 2.1. *Let e and c be positive constants. If n is sufficiently large, and b is a constant sufficiently near c, f_b^n is on the e-neighborhood N_e of f_c^0.*

Suppose the lemma false. There will then exist a constant $e > 0$ and a subsequence F^r of the functions f^n together with points p_r, $r = 1, 2, \cdots$, such that p_r is not on N_e, while

$$c \geqq \sup \lim_{r=\infty} F^r(p_r). \tag{2.3}$$

By virtue of Hypothesis III a subsequence of the points p_r converges to a point q of M.

Without loss of generality we may suppose that $p_1, p_2, \cdots$ is this subsequence. By hypothesis q is not on N_e so that $f^0(q) > c$. But it follows from II that

$$\inf \lim_{r=\infty} F^r(p_r) \geqq f^0(q). \tag{2.4}$$

From (2.3) and (2.4) it follows that $c \geqq f^0(q)$. From this contradiction we infer the truth of the lemma.

HYPOTHESIS IV. *Corresponding to each compact subset A of M, there exists a constant κ^A such that for each integer n, $f^n \leqq \kappa^A f^0$ on A.*

We shall make use of chains and cycles taken in the sense of Vietoris with coefficients in the field of integers mod 2. Cf. M, §1. If p is a point of M, p^* shall represent a 0-cycle each of whose component 0-cycles is on p. This convention will be permanent. The following lemma is an immediate consequence of our definitions.

LEMMA 2.2. *A necessary and sufficient condition that two points p and q belong to the same component of a compact subset A of M is that $p^* + q^* \sim 0$ on A.*

Let f be a function defined on M. A set A will be termed *minimizing* relative to f if f equals a constant c on A, and if there exists an e-neighborhood A_e of A such that $f > C$ on $A_e - A$. If f satisfies Hypothesis I, each minimizing set ω of f is closed. For at a limit point q of ω, $f(q) \leqq c$ by virtue of the lower semi-continuity of f. It follows from the definition of a minimizing set q that q belongs to ω. Hence ω is closed.

A subset of M on which f is bounded will be termed *f-bounded.* We shall prove the following theorem.

THEOREM 2.1. *Let y and z be points respectively of two disjoint minimizing sets ω_1 and ω_2 of f^0. If y and z are connected by an f^0-bounded subset of M and Hypotheses I to IV are satisfied, there exists a subsequence F^r of the functions f^n with the following properties. There exists a least value μ_r of c for which*

$$p^* + q^* \sim 0 \qquad (\text{on } F^r_c),$$

with

$$\mu_r > \max [F^r(y), F^r(z)],$$

while the numbers μ_r converge to a limit μ such that

$$\mu > \max [f^0(y), f^0(z)].$$

We begin the proof of this theorem with certain lemmas. The first lemma concerns a compact subset C of M. A proof may be found in Morse,[4] page 430.

LEMMA 2.3. *If u is a k-cycle on C such that $u \sim 0$ on C_e for each positive e, then $u \sim 0$ on C.*

We shall prove the following lemma.

LEMMA 2.4. *Under the hypotheses of the theorem there exists a least value ν_0 of c such that $y^* + z^* \sim 0$ on f^0_c.*

It follows from the first hypothesis of the theorem and from Lemma 2.2 that there is at least one value of c such that $y^* + z^* \sim 0$ on f^0_c. Let ν_0 be the greatest lower bound of such values of c. There then exist values of c arbitrarily near ν_0 such that $y^* + z^* \sim 0$ on f^0_c. But it follows from Hypothesis I that if c is sufficiently near ν_0, f^0_c lies on an arbitrarily small neighborhood N_e of $f^0_{\nu_0}$. Hence $y^* + z^* \sim 0$ on N_e. We set $C = f^0_{\nu_0}$. We see that $y^* + z^*$ is on C. It follows from Lemma 2.3 that $y^* + z^* \sim 0$ on C, and the proof of Lemma 2.4 is complete.

The points y and z are connected on $f^0_{\nu_0}$. But y and z belong to disjoint minimizing sets. These sets are compact, and hence at a positive distance from each other. It follows that

$$\nu_0 > f^0(y), \qquad \nu_0 > f^0(z). \tag{2.5}$$

The set $f^0_{\nu_0}$ is compact. Hence on this set in accordance with IV, $f^n \leqq \kappa\nu_0$, where κ is a constant. Hence $y^* + z^* \sim 0$ on $f^n_{\kappa\nu_0}$. Let ν_n be the least number c such that $y^* + z^* \sim 0$ on f^n_c. That ν_n exists follows as in the proof of Lemma 2.4. We see that

$$\nu_n \leqq \kappa\nu_0 \qquad (n = 1, 2, \cdots). \tag{2.6}$$

[4] Morse. *Rank and span in functional topology.* Annals of Mathematics, vol. 41 (1940), pp. 419–454.

LEMMA 2.5. *If* ν_n *is the least number* c *such that* $y^* + z^* \sim 0$ *on* f_c^n, $n = 0, 1, \cdots$, *and if* $\mu = \inf \lim_{n\to\infty} \nu_n$, *then*

$$\mu \geqq \nu_0 . \tag{2.7}$$

Let μ_r be a subsequence of the numbers ν_n such that

$$\lim_{r\to\infty} \mu_r = \mu. \tag{2.8}$$

Let F^r be the corresponding subsequence of the functions f^n so that $y^* + z^* \sim 0$ on $F^r_{\mu_r}$. Let e be an arbitrary positive constant, and set $f^0_\mu = C$. It follows from Lemma 2.1 that $F^r_{\mu_r} \subset C_e$, provided r is sufficiently large. Hence $y^* + z^* \sim 0$ on C_e. It follows from Lemma 2.3 that $y^* + z^* \sim 0$ on C. From the definition of ν_0 we infer that $\mu \geqq \nu_0$, and the lemma is proved.

We shall now prove Theorem 2.1, making use of the sequence F^r and the numbers μ_r of the proof of Lemma 2.5. Observe that

$$\mu > \max [f^0(y), f^0(z)]$$

in accordance with (2.5) and (2.7). Recall that F^r and μ_r converge respectively to f^0 and μ. Hence if ρ is a sufficiently large integer

$$\mu_r > \max [F^r(y), F^r(z)] \qquad (r \geqq \rho).$$

Theorem 2.1 is accordingly satisfied by sequences F^r and μ_r, starting with integers $r \geqq \rho$.

3. *h*- and *fh*-critical points. Let F be a function defined on M. We refer to the definition of an F-deformation and a homotopic critical point of F (written h-critical point) as given on page 30 of M. Let $F^1, F^2, \cdots$ be a subsequence of the sequence $f^1, f^2, \cdots$ converging to f^0, and let q_r be an h-critical point of F^r. Any limit point q of the point set q_r will be termed an *fh-critical point* of f^0. This term is relative to the sequence f^n, to which the letter f refers in the term fh-critical point.

LEMMA 3.1. *The set of fh-critical points of* f^0 *is closed, and the set of h-critical points of* f^0 *is closed among points at any given level* μ.

That the set of h-critical points of f^0 is closed at the level μ follows at once from the definition of an h-critical point.

To prove that the set of fh-critical points of f^0 is closed, let p be a limit point of a sequence p_r of fh-critical points of f^0. For each fixed r, p_r is by definition the limit of a sequence $p(r, s)$, $s = 1, 2, \cdots$, of h-critical points of functions $f^{n(r,s)}$ where the integer $n(r, s)$ increases with s. Hence if integers $s_1, s_2, \cdots$ are successively chosen with s_r sufficiently large, the sequence $p(r, s_r)$ will converge to p as r becomes infinite, while the integers $n(r, s_r)$ will increase with r. Hence p is an fh-critical point of f^0, and the proof of the lemma is complete.

To continue we need two additional hypotheses. We refer to MT, page 445, for the definition of weak upper-reducibility.

HYPOTHESIS V. *The functions $f^1, f^2, \cdots$ shall be weakly upper-reducible.*

Lemma 8.1 of M clearly holds if weak upper-reducibility replaces upper-reducibility. We shall use Lemma 8.1 of M in proving the following lemma. We refer to Theorem 2.1 and the constants μ, μ_r and functions F^r described therein.

LEMMA 3.2. *The function F^r of Theorem 2.1 assumes the value μ_r in at least one h-critical point w_r such that y, z, w_r are in the same component of $F^r_{\mu_r}$.*

Let κ be the component of $F^r_{\mu_r}$ which contains y and z. The set κ is compact. Suppose there is no h-critical point of $F^r_{\mu_r}$ on κ at the level μ_r. By virtue of Lemma 8.1 of M there will then exist an F^r-deformation of κ into a set definitely below μ_r. Hence $y^* + z^* \sim 0$ on F^r_c for some $c < \mu_r$, contrary to the definition of μ_r.

We infer the truth of the lemma.

HYPOTHESIS VI. (a). *Let p_r be an h-critical point of f^{n_r}, $n_1 < n_2 < \cdots$, such that p_r converges to p as r becomes infinite. Then $f^{n_r}(p_r)$ shall converge to $f^0(p)$.* (b) *Let p_r be an h-critical point of f^0 such that p_r converges to p. Then $f^0(p_r)$ shall converge to $f^0(p)$.*

A point q will be said to be of *non-minimizing* type relative to $f(p)$ if there is a point p in every neighborhood of q such that $f(p) < f(q)$. With this understood we come to a major theorem.

THEOREM 3.1. *If f^0 admits two disjoint minimizing sets ω_1 and ω_2 and if there are points y and z in ω_1 and ω_2 respectively which are connected by an f^0-bounded set, if moreover Hypotheses I to VI are satisfied, then f^0 possesses at least one fh- or h-critical point of non-minimizing type.*

We refer to the numbers μ_r and μ of Theorem 2.1, and prove the following:

(i). *The set σ of h- and fh-critical points of f^0 at the level μ is closed.*

This follows from Lemma 3.1.

(ii). *The set σ is not empty.*

The h-critical point w_r of Lemma 3.2 is a point at which $F^r(w_r) = \mu_r$. These points w_r have at least one cluster point w by virtue of Hypothesis III. It follows from Hypothesis VI (a) that $f^0(w) = \mu$. Thus w is an fh-critical point belonging to σ, so that σ is not empty.

(iii). *There is at least one point of σ of non-minimizing type.*

The set σ, on which $f^0 = \mu$, fails to contain the points y and z of Theorem 2.1, since μ exceeds $f^0(y)$ and $f^0(z)$ as stated in Theorem 2.1. Observe that y, z, and w lie on the same component of f^0_μ since y, z, and w_r lie on the same component of $F^r_{\mu_r}$, and the maximum distance of points of $F^r_{\mu_r}$ from f^0_μ tends to zero as μ_r tends to μ. The set σ cannot be a minimizing set, otherwise σ would be at a positive distance from its complement on f^0_μ, contrary to the fact that y, z, and w are connected on f^0_μ, while y and z are not in σ and w is in σ. Since σ is compact and not a minimizing set of f^0, there exists at least one point of σ of non-minimizing type.

The proof of the theorem is complete.

4. The application to minimal surface theory. Let g^0 be a simple, closed, rectifiable curve in a space of coordinates $(x_1, \cdots, x_n)$. We suppose g^0 has the vector form

$$\mathbf{x} = \mathbf{g}^0(s),$$

where s is the arc length along g^0. The curve g^0 is given with a sense and an origin $s = 0$. So given, g^0 will be termed *coordinated*. It will be convenient to suppose that the total length of g^0 is 2π. We shall assume that g^0 satisfies the following condition.

Chord arc hypothesis. Under this hypothesis the ratio of the length of an arbitrary chord of g^0 to the length of the minimum subtended arc of g^0 shall be bounded from zero for all chords of g^0.

Any simple, closed, regular curve of class C^1 satisfies this hypothesis, as does a simple closed curve composed of a finite set of regular arcs of class C^1, provided at each corner p the two tangent rays directed from p never make a null angle. In particular any simple closed polygon is admissible. This condition is less restrictive than the corresponding condition of Shiffman (loc. cit.). For under Shiffman's condition two tangent rays at a corner cannot make angles less than a right angle, while it is easy to show that under Shiffman's condition the chord arc hypothesis is always satisfied.

A sequence h^n of closed, rectifiable, coordinated curves $\mathbf{x} = \mathbf{h}^n(s)$ will be said to *converge in length*[5] to h^0: $\mathbf{x} = \mathbf{h}^0(s)$ if

$$\lim_{n=\infty} \mathbf{h}^n(s) = \mathbf{h}^0(s)$$

uniformly on each bounded interval for s. It is clear that there exists a sequence of simple, closed, coordinated polygons which converge in length to g^0. The corners of these polygons can be "rounded off" to obtain a sequence of simple, regular, coordinated, closed curves g^n of class C^2 converging in length to g^0. Without loss of generality we can suppose that the length of each curve g^n is 2π, since if this is not already the case, a magnification from the origin in the ratio ρ_n to 1 will bring this about, where ρ_n must be suitably chosen and so chosen converges to 1. Let $\mathbf{x} = \mathbf{g}^n(s)$ be the vector representation of g^n.

Let $\varphi(\alpha)$ be a continuous non-decreasing function of α such that

$$\varphi(\alpha + 2\pi) \equiv \varphi(\alpha) + 2\pi. \tag{4.1}$$

[5] The coördinated curves h^n converge in length to h^0 if the following three conditions are fulfilled. The sensed curve h^n converges to the sensed curve h^0 according to Fréchet, the length of h^n converges to the length of h^0, and the point $s = 0$ on h^n converges to the point $s = 0$ on h^0. For related, but variant, definitions of convergence in length, see the following three sources. Adams and Lewy. *On convergence in length.* Duke Mathematical Journal, vol. 1 (1935), pp. 19–26. McShane. *Curve space topologies associated with variational problems.* Ann. Scuola Norm. Super. Pisa, (2) vol. 9 (1940), pp. 45–60. Morse. *The calculus of variations in the large.* American Mathematical Society Colloquium Publications (1934), p. 209.

We suppose that $\varphi(\alpha)$ satisfies a *three point condition* defined as follows. The relation $\alpha = \varphi(\alpha)$ shall hold for three given distinct values α_1, α_2, α_3, of α on the interval $0 \leqq \alpha < 2\pi$, where α_1, α_2, α_3 are independent of φ. A function $\varphi(\alpha)$ of this nature will be termed *admissible*. We shall impose another condition on the functions φ when we define the space M.

We admit representations of g^n of the form

$$\mathbf{x} \equiv \mathbf{g}^n[\varphi(\alpha)] \equiv \mathbf{p}^n(\alpha) \qquad (n = 0, 1, \cdots). \tag{4.2}$$

We shall consider the function $\varphi(\alpha)$ as a point in an abstract metric space in which the distance between two points $\varphi(\alpha)$ and $\psi(\alpha)$ shall be the number

$$\psi\varphi = \max | \varphi(\alpha) - \psi(\alpha) | \qquad (0 \leqq \alpha \leqq 2\pi). \tag{4.3}$$

The function $f^n(\varphi)$ shall be defined as the Douglas function[6]

$$f^n(\varphi) = \frac{1}{16\pi} \iint_\omega \frac{[\mathbf{p}^n(\alpha) - \mathbf{p}^n(\beta)]^2}{\sin^2 \dfrac{(\alpha - \beta)}{2}} \, d\alpha \, d\beta, \quad (n = 0, 1, \cdots), \tag{4.4}$$

where ω denotes the parallelogram

$$0 \leqq \beta \leqq 2\pi,$$
$$\beta - \pi \leqq \alpha \leqq \beta + \pi.$$

The Douglas function is an improper integral with an integrand which is singular when $\alpha = \beta$.

As in MT we introduce the improper integral

$$H(\varphi) = \frac{1}{16\pi} \iint_\omega \frac{[\varphi(\alpha) - \varphi(\beta)]^2}{\sin^2 \dfrac{(\alpha - \beta)}{2}} \, d\alpha \, d\beta. \tag{4.5}$$

It follows from the chord arc hypothesis on g^0 and the ordinary properties of length that there exists a constant κ such that

$$[\mathbf{p}^0(\alpha) - \mathbf{p}^0(\beta)]^2 \leqq [\varphi(\alpha) - \varphi(\beta)]^2 \leqq \kappa[\mathbf{p}^0(\alpha) - \mathbf{p}^0(\beta)]^2, \tag{4.6}$$

where κ is independent of the choice of φ. From (4.6) we see that

$$f^0(\varphi) \leqq H(\varphi) \leqq \kappa f^0(\varphi). \tag{4.7}$$

Upon comparing $f^n(\varphi)$ with $H(\varphi)$ we find that

$$f^n(\varphi) \leqq H(\varphi) \leqq \kappa f^0(\varphi). \tag{4.8}$$

Moreover g^n is regular and of class C^2 from which we infer that the ratio of an arbitrary chord length of g^n to either of the corresponding arc lengths exceeds

---

[6] Douglas. I. *Solution of the problem of Plateau.* Transactions of the American Mathematical Society, vol. 33 (1931), pp. 263-321. II. *The mapping theorem of Koebe and the problem of Plateau.* Journal of Mathematics and Physics, vol. 10 (1931), pp. 106-130.

some positive constant κ_n. It follows as in the proof of (4.8) that $f^0 \leqq \kappa_n f^n$. From this result and from (4.8) we conclude that $f^n(\varphi)$ is finite if and only if $f^0(\varphi)$ is finite.

We shall restrict the space M of points φ to points φ for which $f^0(\varphi)$ is finite.

We shall prove the following theorem.

THEOREM 4.1. *At each point of M,*

$$\lim_{n=\infty} f^n(\varphi) = f^0(\varphi). \tag{4.9}$$

Let e be a positive constant less than π and let ω_e be the subset of the parallelogram ω on which $|\alpha - \beta| < e$. We set

$$\omega = \omega_e + \omega_e^*, \qquad (\omega_e \cdot \omega_e^*) = 0,$$

and

$$f^n(\varphi) = F(\varphi, n, e) + F^*(\varphi, n, e), \qquad (n = 0, 1, \cdots), \tag{4.10}$$

where $F(\varphi, n, e)$ and $F^*(\varphi, n, e)$ are defined as is $f^n(\varphi)$ except that ω_e and ω_e^* shall replace ω as the respective domains of integration. From (4.6) we see that

$$F(\varphi, n, e) \leqq \kappa F(\varphi, 0, e). \tag{4.11}$$

By hypothesis $f^0(\varphi)$ is convergent so that $F(\varphi, 0, e)$ is convergent. Noting that the integral $F^*(\varphi, n, e)$ is proper, we see that

$$\lim_{n=\infty} F^*(\varphi, n, e) = F^*(\varphi, 0, e). \tag{4.12}$$

To establish the theorem, observe that

$$|f^n(\varphi) - f^0(\varphi)| \leqq |F(\varphi, n, e) - F(\varphi, 0, e)| + |F^*(\varphi, n, e) - F^*(\varphi, 0, e)|. \tag{4.13}$$

Corresponding to an arbitrary positive constant η let $e > 0$ be chosen so small that $|F(\varphi, 0, e)| < \eta$. This is possible since $F(\varphi, 0, e)$ is a convergent integral. Making use of (4.11) we see that

$$|F(\varphi, n, e) - F(\varphi, 0, e)| \leqq \kappa\eta + \eta.$$

With e so chosen, let n be so large that

$$|F^*(\varphi, n, e) - F^*(\varphi, 0, e)| < \eta.$$

Making use of (4.13), we then find that

$$|f^n(\varphi) - f^0(\varphi)| \leqq \kappa\eta + \eta + \eta,$$

and the theorem follows at once.

5. Verification of hypotheses. We shall show that the Douglas functions $f^n(\varphi)$ defined in the preceding section satisfy Hypotheses I to VI of §§1 to 3.

HYPOTHESIS I. That M is boundedly compact relative to $f^n(\varphi)$ follows from the work of Douglas II. See also Radó[7] V, 17, and Courant.[8]

HYPOTHESIS II. To verify this hypothesis we recall the origin of the Douglas function. Let g be a simple, closed, rectifiable, coordinated curve given with a vector representation $\mathbf{x} = \mathbf{g}(s)$ in terms of arc length s. Let $\varphi(\alpha)$ be an admissible function φ. Corresponding to the representation $\mathbf{x} = \mathbf{p}(\alpha) = \mathbf{g}[\varphi(\alpha)]$ there exists a harmonic surface $\mathbf{x} = \mathbf{x}(u, v)$ defined and continuous for $u^2 + v^2 \leqq 1$ and such that

$$\mathbf{x}\,(\cos\theta, \sin\theta) \equiv \mathbf{p}(\theta),$$

where r and θ are polar coordinates in the (u, v)-plane. We shall say that the harmonic surface $\mathbf{x} = \mathbf{x}(u, v)$ is *defined* by the pair (g, φ).

The Douglas function $f(\varphi)$ corresponding to g equals the Dirichlet sum

$$D(g, \varphi) = \frac{1}{2}\iint_R \left[\left(\frac{\partial \mathbf{x}}{\partial u}\right)^2 + \left(\frac{\partial \mathbf{x}}{\partial v}\right)^2\right] du\,dv,$$

where R represents the region $u^2 + v^2 < 1$.

Let φ^n be a sequence of points on M converging to φ^0. Recall that g^n converges in length to g^0. Hypothesis II is satisfied if

$$\inf_{n=\infty}\lim D(g^n, \varphi^n) \geqq D(g^0, \varphi^0). \tag{5.1}$$

Relation (5.1) is a consequence of the lower semi-continuity of $D(g, \varphi)$ of which a conventional proof may be indicated as follows. Let $D(g, \varphi, r)$ denote the integral obtained from $D(g, \varphi)$ upon replacing R by the region $u^2 + v^2 < r < 1$. By definition

$$D(g, \varphi) = \lim_{r=1} D(g, \varphi, r).$$

The lower semi-continuity of $D(g, \varphi)$ follows from the fact that $D(g, \varphi, r)$ is continuous in (g, φ), positive and non-decreasing in r.

Relation (5.1) holds, and Hypothesis II follows.

HYPOTHESIS III. Let $\varphi^r(\alpha)$ be an infinite sequence of points φ on M such that

$$f^{n_r}(\varphi^r) \leqq c \qquad (n_1 < n_2 < \cdots; r = 1, 2, \cdots)$$

for some constant c and suitable choices of the integers n_r. To establish III we must prove that some subsequence of the sequence φ^r converges to a point on M.

According to the theory of bounded monotone functions there exists a sub-

[7] Radó. *On the problem of Plateau.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Berlin (1933).

[8] Courant. *Plateau's problem and Dirichlet's principle.* Annals of Mathematics, vol. 38 (1937), pp. 679–724.

sequence of the functions φ^r which converges for each value of α to a non-decreasing function $\theta(\alpha)$. The function $\theta(\alpha)$ satisfies the three point condition and the condition

$$\theta(\alpha + 2\pi) \equiv \theta(\alpha) + 2\pi.$$

It remains to show that $\theta(\alpha)$ is continuous.

Whether continuous or not $\theta(\alpha)$ possesses right and left limits a_1 and a_2 at each point α_0. Moreover $| a_1 - a_2 | \neq 2\pi$, since $\theta(\alpha)$ satisfies the three point condition. If $| a_1 - a_2 | \neq 0$, it follows as in the proof of Theorem III of Douglas II, page 113, that

$$\sup \lim_{r=\infty} D(g^{n_r}, \varphi^r) = \infty.$$

But

$$f^{n_r}(\varphi^r) = D(g^{n_r}, \varphi^r) \leqq c.$$

From this contradiction we infer that $a_1 = a_2$ so that $\theta(\alpha)$ is continuous. Thus $\theta(\alpha)$ defines a point of M, and Hypothesis III is verified.

Hypothesis IV. This hypothesis is an immediate consequence of (4.8).

Hypothesis V. The functions $f^n(\varphi)$ are weakly upper-reducible for $n > 0$ in accordance with Theorem 5.1 of MT.

The proof of Hypothesis VI is more involved.

Hypothesis VI. Let g be a simple, closed, coordinated, rectifiable curve, and $\varphi(\alpha)$ a transformation which is admissible in the sense of §4. Let $A(g, \varphi)$ be the area of the harmonic surface defined by the pair (g, φ). In a forthcoming paper we shall prove the following theorem of use in establishing VI.

(i) *The area $A(g, \varphi)$ is continuous in its arguments.*

The notion of continuity of $A(g, \varphi)$ is relative to the concept of convergence of g and φ. Convergence of curves g shall be convergence in length, and convergence of φ, convergence defined by the distance function $\varphi\psi$ of (4.3).

A pair (g, φ) which defines a minimal surface will be called *differentially critical.* When (g, φ) is differentially critical,

$$D(g, \varphi) = A(g, \varphi), \tag{5.2}$$

as is well-known. Cf. Douglas II, Theorem I. When φ is an h-critical point of the function $f(\varphi) = D(g, \varphi)$, (g, φ) will be termed *h-critical.* That an h-critical pair is a differentially critical pair follows from Theorem 6.2 of MT. Hence (5.2) holds for h-critical pairs. From (i) we draw the following conclusion.

(ii). *The function defined by $D(g, \varphi)$ on the subset H of h-critical pairs (g, φ) is continuous on H.*

Hypothesis VI (b) is satisfied by virtue of (ii). To verify Hypothesis VI (a) we return to the sequence $f^n(\varphi)$, and prove the following lemma.

Lemma 5.1. *If ψ^0 is an fh-critical point of f^0, the harmonic surface defined by (g^0, ψ^0) is minimal.*

By definition of an fh-critical point of f^0, ψ^0 is the limit of a sequence of h-critical points ψ^r belonging respectively to a subsequence f^{n_r} of the function f^n. Let

$$S^r\colon \mathbf{x} = \mathbf{x}^r(u, v) \qquad (r = 1, 2, \cdots)$$

be the minimal surface defined by (g^{n_r}, ψ^r), and let

$$S^0\colon \mathbf{x} = \mathbf{x}^0(u, v)$$

be the harmonic surface defined by (g^0, ψ^0). The harmonic surface S^0 will be minimal if its differential coefficients satisfy the conditions

$$E(u, v) \equiv G(u, v), \qquad F(u, v) \equiv 0 \qquad (u^2 + v^2 < 1). \tag{5.3}$$

Let E^r, F^r, G^r denote the corresponding differential coefficients of S^r. Recall that

$$\lim_{r=\infty} (g^{n_r}, \psi^r) = (g^0, \psi^0).$$

Upon representing $\mathbf{x}^r(u, v)$ by means of the Poisson integral, one finds that E^r, F^r, G^r converge to E, F, G as r becomes infinite. But S^r is minimal, so that conditions of the form (5.3) hold for S^r. It follows that (5.3) holds for S^0, so that S^0 is minimal.

The proof of the lemma is complete.

By virtue of the lemma, (5.2) holds on the set of fh-critical points of f^0. Hypothesis VI (a) is satisfied by virtue of (i).

Our principal theorem is as follows.

THEOREM 5.1. *Let g^0 be a simple, closed, rectifiable curve g^0 which satisfies the chord arc hypothesis. If g^0 bounds two minimal surfaces of disc type defined respectively by points on disjoint minimizing sets of the Douglas function $f^0(\varphi)$, then g^0 also bounds a minimal surface of disc type but not of minimum type.*

It follows from the hypotheses of the theorem that $f^0(\varphi)$ possesses two disjoint minimizing subsets ω_1 and ω_2. Let φ and ψ be points on ω_1 and ω_2 respectively. The 1-parameter family of points

$$t\psi(\alpha) + (1 - t)\varphi(\alpha), \qquad (0 \leqq t \leqq 1),$$

is an f^0-bounded arc connecting φ with ψ as has been shown in MT, page 451. The first hypothesis of Theorem 3.1 is accordingly satisfied.

Hypotheses I to VI hold for the sequence $f^n(\varphi)$ as we have seen. Theorem 3.1 thus applies, and we infer the existence of at least one fh- or h-critical point ψ^0 of f^0 not of minimum type. But we have just seen that such a point ψ^0 defines a minimal surface of disc type bounded by g^0; this surface is not of minimum type relative to the Douglas function f^0.

The proof of the theorem is complete.

THE INSTITUTE FOR ADVANCED STUDY AND
PRINCETON UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

ON A LEMMA OF McSHANE[1]

BY TIBOR RADÓ

(Received January 4, 1940)

1. McShane, in his work on the lower semi-continuity of double integrals in Calculus of Variations,[2] established the following important lemma.[3] In the unit square Q: $0 \leqq u \leqq 1$, $0 \leqq v \leqq 1$, let there be given triples of continuous functions $x(u, v)$, $y(u, v)$, $z(u, v)$, $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ such that the following conditions are satisfied.

I. The partial derivatives x_u, x_v, y_u, y_v, z_u, z_v exist almost everywhere in Q.

II. The Jacobians

$$X = y_u z_v - y_v z_u, \qquad Y = z_u x_v - z_v x_u, \qquad Z = x_u y_v - x_v y_u$$

are summable in Q.

III. The functions $x(u, v)$, $y(u, v)$, $z(u, v)$ are absolutely continuous on the boundary B of Q.

IV. We have

$$\iint_Q X\,du\,dv = \frac{1}{2}\int_B (y\,dz - z\,dy), \qquad \iint_Q Y\,du\,dv = \frac{1}{2}\int_B (z\,dx - x\,dz),$$

$$\iint_Q Z\,du\,dv = \frac{1}{2}\int_B (x\,dy - y\,dx).$$

V. The functions $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ are quasi-linear[4] in Q.

VI. $x_n(u, v) \to x(u, v)$, $y_n(u, v) \to y(u, v)$, $z_n(u, v) \to z(u, v)$ uniformly in Q.

Then for every choice of the constants a, b, c there exists a sequence $\{V_n\}$ *of measurable subsets of Q such that*

$$\iint_{V_n} (aX_n + bY_n + cZ_n)\,du\,dv \xrightarrow[n\to\infty]{} \iint_Q (aX + bY + cZ)\,du\,dv,$$

where

$$X_n = y_{nu}z_{nv} - y_{nv}z_{nu}, \quad Y_n = z_{nu}x_{nv} - z_{nv}x_{nu}, \quad Z_n = x_{nu}y_{nv} - x_{nv}y_{nu}.$$

2. In view of the importance of this lemma, it seemed to be of interest to investigate the possibility of replacing the conditions I–IV of McShane, con-

[1] Presented to the American Mathematical Society at the meeting in Chicago, April 1940.

[2] E. J. McShane, *Integrals over surfaces in parametric form*, Annals of Mathematics, vol. 34, 1933, pp. 815–838.

[3] Loc. cit.,[2] p. 829.

[4] A continuous function $f(u, v)$ is quasi-linear in Q if Q can be subdivided into a finite number of triangles on each of which $f(u, v)$ is linear.

cerned with the limit triple $x(u, v)$, $y(u, v)$, $z(u, v)$, by less restrictive ones. It should be noted that the purpose of conditions III and IV is to provide, for the integral of $aX + bY + cZ$, a geometrical interpretation which is essential for the proof of McShane. It may be therefore of interest to point out that *the lemma remains valid if the apparently decisive conditions* III *and* IV *are dropped.* The purpose of this paper is to prove this assertion. Applications of the result will be considered on another occasion.

3. As stated above, the lemma involves three arbitrary constants a, b, c, but it is evident from the work of McShane that it is sufficient to establish the apparently very special case $a = b = 0$, $c = 1$, since the general case can then be disposed of by a very simple device. For this reason, and also for easier reference, we wish to state our result for this special case explicitly.

4. THEOREM. *In the unit square* Q: $0 \leqq u \leqq 1$, $0 \leqq v \leqq 1$, *let there be given pairs of functions* $x(u, v)$, $y(u, v)$, $x_n(u, v)$, $y_n(u, v)$ *such that the following conditions are satisfied.*

I. *The partial derivatives* x_u, x_v, y_u, y_v *exist almost everywhere in* Q.

II. *The Jacobian* $J = x_u y_v - x_v y_u$ *is summable in* Q.

III. *The functions* $x_n(u, v)$, $y_n(u, v)$ *are quasi-linear in* Q.

IV. $x_n(u, v) \to x(u, v)$, $y_n(u, v) \to y(u, v)$ *uniformly in* Q.

Then there exists a sequence of measurable subsets V_n *of* Q *such that*

$$\iint_{V_n} J_n \, du \, dv \to \iint_Q J \, du \, dv,$$

where $J_n = x_{nu} y_{nv} - x_{nv} y_{nu}$.

5. Applying a device due to McShane,[5] we proceed as follows to derive from this statement the one described in section 2. Let us consider the triples $x(u, v)$, $y(u, v)$, $z(u, v)$, $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$, but let us assume only that conditions I, II, V, VI of section 1 are satisfied. Let there be given a set of constants a, b, c. Without loss of generality we can assume that $a^2 + b^2 + c^2 = 1$. We can find then an orthogonal matrix, with determinant $+1$, of the form

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a & b & c \end{pmatrix}.$$

Let us put

$$\bar{x}(u, v) = a_{11}x(u, v) + a_{12}y(u, v) + a_{13}z(u, v),$$
$$\bar{x}_n(u, v) = a_{11}x_n(u, v) + a_{12}y_n(u, v) + a_{13}z_n(u, v),$$
$$\bar{y}(u, v) = a_{21}x(u, v) + a_{22}y(u, v) + a_{23}z(u, v),$$
$$\bar{y}_n(u, v) = a_{21}x_n(u, v) + a_{22}y_n(u, v) + a_{23}z_n(u, v).$$

Let $\bar{J}$, $\bar{J}_n$ denote the Jacobians of the pairs $\bar{x}$, $\bar{y}$ and $\bar{x}_n$, $\bar{y}_n$ respectively. Clearly, the pairs $\bar{x}$, $\bar{y}$, $\bar{x}_n$, $\bar{y}_n$ satisfy the conditions of section 4 and therefore we have,

[5] See loc. cit.[3]

by the theorem stated in that section, a sequence $\{V_n\}$ of measurable subsets of Q such that

$$\iint_{V_n} \mathfrak{J}_n \, du\, dv \to \iint_{Q} \mathfrak{J}\, du\, dv.$$

This sequence possesses the desired property, since we have by direct computation

$$\mathfrak{J} = aX + bY + cZ, \qquad \mathfrak{J}_n = aX_n + bY_n + cZ_n .$$

6. The rest of this paper contains the proof of the theorem of section 4. Sections 7–12 contain some facts needed in the sequel, while sections 13–17 contain the proof itself.

7. Let G be a bounded measurable set in the xy plane and $h(x, y)$, $h_n(x, y)$ summable functions on G such that $h_n(x, y) \to h(x, y)$ on G. We assume that these functions do not take on the values $\pm\infty$. Under these conditions, there exists a sequence of measurable subsets G_n of G such that

$$\iint_{G_n} h_n \, dx\, dy \to \iint_{G} h\, dx\, dy.$$

PROOF. Define

$$\kappa_n = \text{gr.l.b.} \left| \iint_{G} h\, dx\, dy - \iint_{E} h_n\, dx\, dy \right|,$$

where the greatest lower bound is taken with respect to all measurable subsets E of G. Give any $\epsilon > 0$. Since $h(x, y)$ is summable on G, we have then an $\eta = \eta(\epsilon) > 0$ such that if S is any measurable subset of G we have

$$\left| \iint_{S} h\, dx\, dy \right| < \epsilon \quad \text{if} \quad |S| < \eta.$$[6]

Since $h_n \to h$ on G, we have by the theorem of Egoroff[7] a measurable subset H of G such that

$$|G - H| < \eta$$

and such that $h_n \to h$ *uniformly* on H. Hence

$$\iint_{H} |h - h_n|\, dx\, dy \to 0 \quad \text{for} \quad n \to \infty,$$

and consequently

$$0 \leqq \overline{\lim_{n\to\infty}}\, \kappa_n \leqq \epsilon,$$

[6] If S is a measurable set, then $|S|$ denotes its Lebesgue measure.

[7] See for instance Saks, *Theory of the integral*, Warszawa 1937, p. 18.

since

$$0 \leqq \kappa_n \leqq \left| \iint_G h\,dx\,dy - \iint_H h_n\,dx\,dy \right|$$

$$\leqq \left| \iint_{G-H} h\,dx\,dy \right| + \iint_H |h - h_n|\,dx\,dy < \epsilon + \iint_H |h - h_n|\,dx\,dy.$$

As $\epsilon > 0$ was arbitrary, it follows that $\kappa_n \to 0$. By the definition of κ_n, we have a measurable subset G_n of G such that

$$\left| \iint_G h\,dx\,dy - \iint_{G_n} h_n\,dx\,dy \right| < \kappa_n + \frac{1}{n}.$$

As $\kappa_n \to 0$, the sequence G_n has therefore the desired property.

8. Suppose the functions $f(u, v)$, $g(u, v)$ are merely continuous on the unit square Q: $0 \leqq u \leqq 1, 0 \leqq v \leqq 1$. Let (u_0, v_0) be an interior point of Q and q a square in Q with center at (u_0, v_0) and sides parallel to the axes. Denote by b the boundary of q. Take four constants $\alpha, \beta, \gamma, \delta$ and put

$$\rho(u_0, v_0, q, f, \alpha, \beta) = \max_{(u,v)\epsilon b} \frac{|f(u, v) - f(u_0, v_0) - \alpha(u - u_0) - \beta(v - v_0)|}{[(u - u_0)^2 + (v - v_0)^2]^{\frac{1}{2}}},$$

$$\rho(u_0, v_0, q, g, \gamma, \delta) = \max_{(u,v)\epsilon b} \frac{|g(u, v) - g(u_0, v_0) - \gamma(u - u_0) - \delta(v - v_0)|}{[(u - u_0)^2 + (v - v_0)^2]^{\frac{1}{2}}}.$$

If there exists *some* sequence of squares q_n such that simultaneously

$$\rho(u_0, v_0, q_n, f, \alpha, \beta) \to 0, \quad \rho(u_0, v_0, q_n, g, \gamma, \delta) \to 0, \quad |q_n| \to 0, \tag{1}$$

then we shall say that the pair $f(u, v)$, $g(u, v)$ satisfies condition C at the point (u_0, v_0) with the constants $\alpha, \beta, \gamma, \delta$.

9. The reader will have noticed that the preceding condition C is merely a greatly weakened form of *total differentiability*. As it might be expected, the methods developed by Rademacher, Stepanoff and others to obtain results of surprising generality concerning total differentiability,[8] can be easily adapted to our condition C. One statement obtained in this manner is as follows.[9] Suppose that the functions $f(u, v)$, $g(u, v)$ are continuous in Q and that the partial derivatives f_u, f_v, g_u, g_v exist almost everywhere in Q. Then condition C is satisfied at almost every point (u_0, v_0) of Q with the constants $\alpha = f_u(u_0, v_0)$, $\beta = f_v(u_0, v_0)$, $\gamma = g_u(u_0, v_0)$, $\delta = g_v(u_0, v_0)$.

10. Consider again a pair of functions $f(u, v)$, $g(u, v)$ which we assume to be merely continuous in Q. These functions define a continuous transformation

[8] See the excellent presentation loc. cit.,[7] Ch. 9.

[9] For the details of the proof, see the author's paper, *On absolutely continuous transformations in the plane*, Duke Math. Journal, vol. 4, 1938, pp. 189–221, in particular p. 219.

$$T: \begin{cases} x = f(u, v) \\ y = g(u, v) \end{cases} \qquad (u, v) \in Q.$$

Let q be a square in Q, with sides parallel to the axes, and let b be the boundary of q. If the point (u, v) describes b in the counter-clockwise sense, its image under T describes in the xy-plane a continuous closed oriented curve b'. In the xy-plane we define a function $\mu(x, y, q, T)$ as follows. If the point (x, y) is on b', then $\mu(x, y, q, T) = 0$. Otherwise $\mu(x, y, q, T)$ is equal to the topological index of (x, y) with respect to b'. We shall denote by $S_1(q, T)$, $S_2(q, T)$ the point-sets in the xy-plane where $\mu(x, y, q, T)$ is equal to $+1$ and -1 respectively. These sets are bounded and open (possibly vacuous).

11. Keeping the notations of the preceding section, let us consider an interior point (u_0, v_0) of Q where the condition C of section 8 is satisfied with some constants $\alpha, \beta, \gamma, \delta$. Let $\{q_n\}$ be a sequence of squares, as specified in section 8, such that the relations (1) hold. We consider the auxiliary transformation

$$T^*: \begin{cases} x = f(u_0, v_0) + \alpha(u - u_0) + \beta(v - v_0), \\ y = g(u_0, v_0) + \gamma(u - u_0) + \delta(v - v_0), \end{cases} \qquad (u, v) \in Q.$$

Assuming that $\alpha\delta - \beta\gamma \neq 0$, the image of the square q_n under T^* is a parallelogram π_n^* in the xy-plane. Suppose first that $\alpha\delta - \beta\gamma > 0$. Then the set $S_1(q_n, T^*)$ is simply the interior of π_n^*, and condition C implies, by a familiar elementary reasoning,[10] that for large values of n the symmetric difference[11] of the sets $S_1(q_n, T^*)$ and $S_1(q_n, T)$ is contained in a narrow strip σ_n, bounded by two parallelograms similar to π_n^*, such that $|\sigma_n|/|\pi_n^*| \to 0$. As $|\pi_n^*|/|q_n| \to \alpha\delta - \beta\gamma$, we also have $|\sigma_n|/|q_n| \to 0$. Thus a fortiori

$$\frac{|S_1(q_n, T)| - |S_1(q_n, T^*)|}{|q_n|} \to 0.$$

But, since T^* is an affine transformation,

$$\frac{|S_1(q_n, T^*)|}{|q_n|} = \frac{|\pi_n^*|}{|q_n|} \to \alpha\delta - \beta\gamma.$$

Hence

$$\frac{|S_1(q_n, T)|}{|q_n|} \to \alpha\delta - \beta\gamma \quad \text{if} \quad \alpha\delta - \beta\gamma > 0.$$

Similarly it follows that

$$-\frac{|S_2(q_n, T)|}{|q_n|} \to \alpha\delta - \beta\gamma \quad \text{if} \quad \alpha\delta - \beta\gamma < 0.$$

---

[10] For a detailed discussion of a practically identical situation, see the author's paper, *Über das Flächenmass rektifizierbarer Flächen*, Math. Annalen, vol. 100, 1928, pp. 445–479, in particular pp. 461–466.

[11] The symmetric difference of two sets A and B consists of those points which belong to exactly one of A, B.

12. Using the notations of section 10, let us assume that $f(u, v)$, $g(u, v)$ are quasi-linear in Q. Let us denote by $J(u, v)$ the Jacobian $f_u g_v - f_v g_u$. Let us take, in the xy-plane, any summable function $F(x, y)$ which takes on only finite values (summability meaning that $F(x, y)$ is summable on every bounded measurable set). Then we have for every square q in Q the transformation formula

$$\iint_q F[f(u, v,), g(u, v)]J(u, v)\, du\, dv = \iint F(x, y)\mu(x, y, q, T)\, dx\, dy.$$[12]

As a matter of fact, this formula holds under very general conditions.[13] For quasi-linear transformations, of course, the formula is practically trivial.

13. We proceed presently to prove the theorem of section 4. Through the rest of the paper the notations of that section will be used. Let q be a square in Q with sides parallel to the axes. We define

$$\lambda_n(q) = \underset{E \subset q}{\text{gr.l.b.}} \left| \iint_q J\, du\, dv - \iint_E J_n\, du\, dv \right|;$$

that is, the greatest lower bound is taken with respect to all measurable subsets E of q. We define further

$$\lambda(q) = \overline{\lim_{n \to \infty}}\, \lambda_n(q).$$

By taking in q a set E of measure zero, we see that

$$\lambda_n(q) \leqq \left| \iint_q J\, du\, dv \right| \leqq \iint_q |J|\, du\, dv, \tag{2}$$

and hence also

$$\lambda(q) \leqq \iint |J|\, du\, dv.$$

Take now in Q any finite or infinite sequence of non-overlapping squares $q_1, q_2, \cdots, q_j, \cdots$, such that

$$|Q - \sum q_j| = 0.$$

We assert that

$$\lambda(Q) \leqq \sum \lambda(q_j).$$

---

[12] As $\mu(x, y, q, T)$ vanishes outside of a sufficiently large circle, the range of integration on the right can be taken as the whole xy-plane.

[13] See, also for further literature, loc. cit.[9] It is worth noting that the quasi-linear character of $f(u, v)$, $g(u, v)$ is used only to secure the above transformation formula. This remark suggests further generalizations which we do not wish to discuss at this time.

PROOF.[14] Observe first that the series on the right is convergent since it is dominated by the series

$$\sum \iint_{q_j} |J|\, du\, dv,$$

which converges because the squares q_j do not overlap. Give now $\epsilon > 0$. By the definition of $\lambda_n(q_j)$, we have a measurable subset E_n^j of q_j such that

$$\left| \iint_{q_j} J\, du\, dv - \iint_{E_n^j} J_n\, du\, dv \right| < \lambda_n(q_j) + \frac{\epsilon}{2^j}. \tag{3}$$

Put

$$E_n = \sum_j E_n^j.$$

Since the squares q_j do not overlap, we have by (3)

$$\lambda_n(Q) \leqq \left| \iint_{Q} J\, du\, dv - \iint_{E_n} J_n\, du\, dv \right|$$
$$\leqq \sum_j \left| \iint_{q_j} J\, du\, dv - \iint_{E_n^j} J_n\, du\, dv \right| < \sum_j \lambda_n(q_j) + \epsilon. \tag{4}$$

Take any positive integer i. We have then by (2)

$$\sum_j \lambda_n(q_j) \leqq \sum_{j=1}^{i} \lambda_n(q_j) + \sum_{j=i+1}^{\infty} \iint_{q_j} |J|\, du\, dv.$$

Hence

$$\overline{\lim_{n\to\infty}} \sum_j \lambda_n(q_j) \leqq \sum_{j=1}^{i} \lambda(q_j) + \sum_{j=i+1}^{\infty} \iint_{q_j} |J|\, du\, dv.$$

Since the squares q_j do not overlap, the second summation on the right converges to zero for $i \to \infty$. Thus

$$\overline{\lim_{n\to\infty}} \sum_j \lambda_n(q_j) \leqq \sum_j \lambda(q_j).$$

From (4) we infer now

$$\lambda(Q) = \overline{\lim_{n\to\infty}}\, \lambda_n(Q) \leqq \sum \lambda(q_j) + \epsilon.$$

Since ϵ was arbitrary, this proves our assertion.

14. We consider now the transformation

$$T: \begin{cases} x = x(u, v), \\ y = y(u, v), \end{cases} \qquad (u, v) \in Q.$$

[14] We discuss the case of an infinite sequence q_j, since the finite case is trivial.

Let q be a square in Q with sides parallel to the axes. We assert the inequality[15]

$$\lambda(q) \leqq \min \left\{ \left| \iint_q J\,du\,dv - |S_1(q, T)| \right|, \left| \iint_q J\,du\,dv + |S_2(q, T)| \right| \right\}, \tag{5}$$

where the sets $S_1(q, T)$, $S_2(q, T)$ are defined as in section 10.

Proof. Let us consider the transformations

$$T_n: \begin{cases} x = x_n(u, v), \\ y = y_n(u, v), \end{cases} \qquad (u, v) \in Q.$$

In the xy-plane, let S be the set where $\mu(x, y, q, T) \neq 0$. Since $x_n(u, v) \to x(u, v)$, $y_n(u, v) \to y(u, v)$ uniformly in Q, and hence in q, we have

$$\mu(x, y, q, T_n) \to \mu(x, y, q, T)$$

on S, and hence also on the sets $S_1(q, T)$, $S_2(q, T)$. Since T_n is quasi-linear, $\mu(x, y, q, T_n)$ is summable. We cannot assert that $\mu(x, y, q, T)$ is summable, but this function is surely summable on $S_1(q, T)$ and $S_2(q, T)$, since $|\mu(x, y, q, T)| = 1$ on these sets. Hence we can apply the remark in section 7 with $h(x, y) = \mu(x, y, q, T)$, $h_n(x, y) = \mu(x, y, q, T_n)$ and first with $G = S_1(q, T)$ and second with $G = S_2(q, T)$. It follows that we have sequences of bounded measurable sets G_n', G_n'' such that

$$\iint_{G_n'} \mu(x, y, q, T_n)\,dx\,dy \to \iint_{S_1(q,T)} \mu(x, y, q, T)\,dx\,dy = |S_1(q, T)|, \tag{6}$$

$$\iint_{G_n''} \mu(x, y, q, T_n)\,dx\,dy \to \iint_{S_2(q,T)} \mu(x, y, q, T)\,dx\,dy = -|S_2(q, T)|.$$

Now let E_n' be the complete model[16] in q of the set G_n' under the transformation T_n. The transformation formula of section 12 yields then, if $F(x, y)$ is taken as the characteristic function of the set G_n', the formula

$$\iint_{G_n'} \mu(x, y, q, T_n)\,dx\,dy = \iint_{E_n'} J_n\,du\,dv. \tag{7}$$

Similarly, if E_n'' is the complete model of G_n'' in q under the transformation T_n, we have

$$\iint_{G_n''} \mu(x, y, q, T_n)\,dx\,dy = \iint_{E_n''} J_n\,du\,dv.$$

Now, by the definition of $\lambda_n(q)$,

$$\lambda_n(q) \leqq \left| \iint_q J\,du\,dv - \iint_{E_n'} J_n\,du\,dv \right|,$$

[15] If a, b are real numbers, then min (a, b) denotes the smaller one of a, b if $a \neq b$ and the common value of a, b if $a = b$.

[16] That is, E_n' is the set of all those points in q whose image under T_n is comprised in G_n'.

and for $n \to \infty$ it follows by (6), (7) that

$$\lambda(q) \leqq \left| \iint_q J\,du\,dv - |S_1(q, T)| \right|.$$

Similarly

$$\lambda(q) \leqq \left| \iint_q J\,du\,dv + |S_2(q, T)| \right|.$$

The last two inequalities imply (5).

15. Denote now by E the set of those interior points (u_0, v_0) of Q where the following conditions are satisfied.

I. x_u, x_v, y_u, y_v exist at (u_0, v_0).

II. Condition C of section 8 is satisfied at (u_0, v_0) with the constants $\alpha = x_u(u_0, v_0)$, $\beta = x_v(u_0, v_0)$, $\gamma = y_u(u_0, v_0)$, $\delta = y_v(u_0, v_0)$.

III. If q is a square with sides parallel to the axes and with center at (u_0, v_0), then

$$\frac{\iint_q J\,du\,dv}{|q|} \to J(u_0, v_0) \quad \text{for} \quad |q| \to 0.$$

Since $J(u, v)$ exists almost everywhere in Q and is summable there by assumption, III is satisfied almost everywhere in Q by a well-known theorem, while II is satisfied almost everywhere in Q by section 9. Thus

$$|Q - E| = 0.$$

16. Now let (u_0, v_0) be a point of the set E of the preceding section. By condition II, we have then a sequence $q_n(u_0, v_0)$ of squares with sides parallel to the axes and with center at (u_0, v_0) for which the relations (1) of section 8 hold with $f(u, v) = x(u, v)$, $g(u, v) = y(u, v)$, $\alpha = x_u(u_0, v_0)$, $\beta = x_v(u_0, v_0)$, $\gamma = y_u(u_0, v_0)$, $\delta = y_v(u_0, v_0)$. By section 11 we have therefore

$$\frac{|S_1(q_n(u_0, v_0), T)|}{|q_n(u_0, v_0)|} \to J(u_0, v_0) \quad \text{if} \quad J(u_0, v_0) > 0,$$

and

$$-\frac{|S_2(q_n(u_0, v_0), T)|}{|q_n(u_0, v_0)|} \to J(u_0, v_0) \quad \text{if} \quad J(u_0, v_0) < 0.$$

But, by condition III of section 15, we also have

$$\frac{\iint_{q_n(u_0, v_0)} J\,du\,dv}{|q_n(u_0, v_0)|} \to J(u_0, v_0).$$

Hence

$$\frac{\displaystyle\iint\limits_{q_n(u_0,v_0)} J\,du\,dv - |S_1(q_n(u_0,v_0),T)|}{|q_n(u_0,v_0)|} \to 0 \quad \text{if} \quad J(u_0,v_0) > 0, \tag{8}$$

and

$$\frac{\displaystyle\iint\limits_{q_n(u_0,v_0)} J\,du\,dv + |S_2(q_n(u_0,v_0),T)|}{|q_n(u_0,v_0)|} \to 0 \quad \text{if} \quad J(u_0,v_0) < 0. \tag{9}$$

By (5) in section 14 the relations (8), (9) imply that

$$\frac{\lambda(q_n(u_0,v_0))}{|q_n(u_0,v_0)|} \to 0$$

if $J(u_0, v_0) \neq 0$. If $J(u_0, v_0) = 0$, then we infer by section 13 and by condition III of section 15 that

$$0 \leqq \frac{\lambda(q_n(u_0,v_0))}{|q_n(u_0,v_0)|} \leqq \frac{\left|\displaystyle\iint\limits_{q_n(u_0,v_0)} J\,du\,dv\right|}{|q_n(u_0,v_0)|} \to |J(u_0,v_0)| = 0.$$

17. Thus we see that for every point (u_0, v_0) of the set E of section 15 there exists a sequence of squares $q_n(u_0, v_0)$, with sides parallel to the axes and with center at (u_0, v_0), such that

$$\frac{\lambda(q_n(u_0,v_0))}{|q_n(u_0,v_0)|} \to 0 \quad \text{and} \quad |q_n(u_0,v_0)| \to 0.$$

Besides, all the squares $q_n(u_0, v_0)$ are comprised in Q. Give now $\epsilon > 0$. If we discard, for each point (u_0, v_0) of E, a finite (sufficiently large) number of the squares $q_n(u_0, v_0)$, the remaining ones will satisfy the relations

$$\lambda(q_n(u_0,v_0)) < \epsilon\,|q_n(u_0,v_0)|, \qquad |q_n(u_0,v_0)| \to 0, \qquad q_n(u_0,v_0) \subset Q.$$

The squares $q_n(u_0, v_0)$ cover the set E in the manner required by the Vitali covering theorem. We can select therefore from amongst the squares $q_n(u_0, v_0)$ a sequence of non-overlapping squares $q_1, q_2, \cdots, q_j, \cdots$ which cover E with the possible exception of a set of measure zero. Since

$$\sum q_j \subset Q \qquad \text{and} \qquad |Q - E| = 0,$$

we also have

$$|Q - \sum q_j| = 0.$$

Hence by section 13

$$\lambda(Q) \leqq \sum \lambda(q_j) \leqq \epsilon \sum |q_j| = \epsilon\,|Q| = \epsilon.$$

As ϵ was arbitrary, it follows that $\lambda(Q) = 0$. By section 13 this implies that

$$\lambda_n(Q) \to 0. \tag{10}$$

By the definition of $\lambda_n(Q)$, there exists a measurable subset V_n of Q such that

$$\left| \iint_Q J \, du \, dv - \iint_{V_n} J_n \, du \, dv \right| < \lambda_n(Q) + \frac{1}{n}.$$

(10) and (11) imply that

$$\iint_{V_n} J_n \, du \, dv \to \iint_Q J \, du \, dv,$$

and thus the theorem of section 4 is proved.

18. As far as applications are concerned, it seems that in some cases it would be sufficient to know that we have a sequence V_n such that

$$\underline{\lim} \iint_{V_n} J_n \, du \, dv \geqq \iint_Q J \, du \, dv.$$

It is however unlikely that this relation can be derived under assumptions weaker than those needed to obtain the precise conclusion of the theorem of section 4.

THE OHIO STATE UNIVERSITY.

Annals of Mathematics
Vol. 42, No. 1, January, 1941

ON FINSLER AND CARTAN GEOMETRIES. III

TWO-DIMENSIONAL FINSLER SPACES WITH RECTILINEAR EXTREMALS[1]

By L. Berwald

(Received August 26, 1939)

Introduction. The principal part of the present paper is devoted to the problem, proposed by P. Funk,[2] of characterizing, in an invariant manner, the two-dimensional Finsler spaces the extremals of which can be given, in a suitable coordinate system, by linear equations. We call such spaces Finsler spaces with rectilinear extremals.

In an introductory section, we explain briefly, from the beginning, the theory of the two-dimensional Finsler spaces developed especially by the author[3] and by E. Cartan.[4] The standpoint of this exposition is predominantly formal. Our aim is to develop Cartan's theory of the two-dimensional Finsler spaces independently of the general theory of equivalence, and to connect it with his later theory of the n-dimensional Finsler spaces.[5] The single new feature in this section is the connection between the tensor FG^{i}_{jkl} and the main scalar of a two-dimensional Finsler space, given in §8.

Section II develops two different methods which lead to an invariant characterization of the two-dimensional Finsler spaces with rectilinear extremals. The first is purely analytical, and is based upon the discussion of the conditions of integrability of a certain system of partial differential equations (§§9–11). The scope of the second method is first to establish necessary and sufficient conditions in order that a two-dimensional general geometry of paths may have rectilinear paths,[6] and then to apply them to a Finsler space. This method has the advantage of showing what is the independent significance of each of the two conditions we obtain (§§12, 13).

In the third section we determine all two-dimensional Finsler spaces with rectilinear extremals the main scalar of which is a function of position only. First we establish some theorems showing that for such a space the main scalar

[1] The first two papers of this series are L. Berwald, [2], [3]. (Numbers in cornered brackets refer to the bibliography at the end of the paper).

[2] See P. Funk, [11].

[3] L. Berwald, [1].

[4] E. Cartan, [5].

[5] E. Cartan, [6].

[6] For an n-dimensional general geometry of paths ($n > 2$), J. Douglas, [7], established the corresponding conditions. When $n = 2$, only one of the two conditions for rectilinear paths results from Douglas' corresponding condition by particularization, the other does not. The reason is that, when $n = 2$, the generalized Weyl projective curvature tensor (H. Weyl, [17]; J. Douglas, [7]) vanishes identically.

is always, and the curvature nearly always, a constant (being necessarily null if the main scalar does not vanish). There exists but one exception: the case in which the main scalar has the value $\pm 3/\sqrt{2}$ (§15). This exceptional case is studied in §§16, 17. In toto, there are six types of two-dimensional Finsler spaces of the desired kind. Four of these types depend on arbitrary constants, a fifth on an arbitrary function of a single variable (§18). Finally we determine all *Landsberg spaces* with rectilinear extremals (§19).

I. Two-dimensional Finsler Spaces

1. The metric. We start with an n-dimensional manifold with coordinates x^i in which a variational problem with the fundamental integral

$$(1.1)\quad s = \int_{t_1}^{t} F(x^1, x^2, \cdots, x^n; x'^1, x'^2, \cdots, x'^n)\, dt = \int_{t_1}^{t} F(x, x')\, dt, \quad \left(x'^i = \frac{dx^i}{dt}\right)$$

is given. We suppose that F is analytic in a certain region $\mathfrak{B}$ of its $2n$ arguments to which we restrict ourselves, and that $\mathfrak{B}$ contains no points with $x'^1 = x'^2 = \cdots = x'^n = 0$. Further F is supposed to be positive and positively homogeneous of the first degree in x'. Finally, we suppose $F_1 > 0$, where

$$(1.2)\qquad F_1 = -\frac{1}{F^2}\,\mathrm{det.}\left(\begin{array}{c|c} \dfrac{\partial^2 F}{\partial x'^i\, \partial x'^k} & \dfrac{\partial F}{\partial x'^k} \\ \hline \dfrac{\partial F}{\partial x'^i} & 0 \end{array}\right).$$

The second factor on the right of (1.2) stands for the determinant of an $(n+1)$-rowed matrix. A manifold of the considered kind is called an *n-dimensional Finsler space*,[7] and F its *fundamental function.*

We interpret s as arc-length of the curve $x^i = x^i(t)$. As element of space we regard the oriented line-element (x, x'), that is, a point (x) and a direction $(\rho x';\ \rho > 0)$ issuing from it. The quantities with which we have to deal (tensors, densities, and so on) depend exclusively on the line-element, i.e. *they are positively homogeneous of degree zero in x'.*

The metric of a Finsler space is based upon the symmetric covariant tensor of the second order

$$(1.3)\qquad g_{ik}(x, x') = \frac{1}{2}\frac{\partial^2(F^2)}{\partial x'^i\, \partial x'^k}.$$

On account of

$$(1.4)\qquad g = \mathrm{det.}\ (g_{ik}) = F^{n+1} F_1 > 0,$$

an n-dimensional euclidean metric is associated with each line-element (x, x') by

$$(1.5)\qquad d\sigma^2 = g_{ik}(x, x')\, dx^i\, dx^k.^{8}$

[7] P. Finsler, [8].

[8] Repeated indices indicate summation.

This metric is particularly used for measuring the length of vectors. According to the homogeneity of F we have

$$g_{ik}x'^i x'^k = F^2, \tag{1.6}$$

$$g_{ik}x'^k = F\frac{\partial F}{\partial x'^i}. \tag{1.7}$$

Therefore orthogonality to the line-element (x, x') in the metric (1.5) is identical with transversality to the line-element.

In consequence of (1.4), the tensor g^{ik} conjugate to g_{ik} exists. By means of g_{ik} and g^{ik} the lowering and raising of indices is defined in the usual manner. We can therefore speak of covariant, contravariant, mixed components of a tensor.

The unit vector of the line-element (x, x') has the contravariant components

$$l^i = \frac{dx^i}{ds} = \frac{x'^i}{F}, \tag{1.8}$$

and the covariant components

$$l_i = \frac{\partial F}{\partial x'^i}. \tag{1.9}$$

Inner multiplication with the unit vector (l) is indicated by a zero, the position of which (above or below) does not matter; for instance

$$T_{0i} = T_{hi}l^h = T^h_{\cdot i}l_h, \qquad T^{i0} = T^{ih}l_h = T^i_{\cdot h}l^h. \tag{1.10}$$

2. Extremals and parallel displacement. The derivative Φ_s. The extremals of the variational problem $\delta s = 0$ have the differential equations

$$\frac{d}{dt}\frac{\partial F}{\partial x'^i} - \frac{\partial F}{\partial x^i} = 0. \tag{2.1}$$

With regard to $F\left(x, \frac{dx}{ds}\right) = 1$ and to the homogeneity of F, (2.1) can be written as

$$\frac{d^2x^i}{ds^2} + 2G^i\left(x, \frac{dx}{ds}\right) = 0 \tag{2.2}$$

or

$$\frac{dl^i}{ds} + 2\frac{G^i(x, x')}{F^2(x, x')} = 0, \tag{2.3}$$

where

$$G^i(x, x') = \tfrac{1}{4}g^{ih}\left[\frac{\partial^2(F^2)}{\partial x'^h\,\partial x^m}x'^m - \frac{\partial(F^2)}{\partial x^h}\right]. \tag{2.4}$$

G^i is therefore positively homogeneous of the second degree in x'.

We indicate by Φ_s the derivative of any function $\Phi(x, x')$ with respect to the arc s of the extremal issuing from the line-element (x, x'). Because of $\dfrac{dF}{ds} = 0$ and of the homogeneity of G^i, we have

$$\Phi_s = \Phi_{(i)} l^i = \Phi_{(0)}, \tag{2.5}$$

where

$$\Phi_{(i)} = \frac{\partial \Phi}{\partial x^i} - \frac{\partial \Phi}{\partial x'^r} \frac{\partial G^r}{\partial x'^i}. \tag{2.6}$$

If Φ does not depend on x', $\Phi_{(i)}$ reduces to the partial derivative $\dfrac{\partial \Phi}{\partial x^i}$. For a scalar Φ, $\Phi_{(i)}$ are the covariant components of a vector. If Φ is a tensor which is not a scalar, the differentiation $\cdots_{(i)}$ does not generate a tensor. In particular we have

$$F_{(i)} = 0. \tag{2.7}$$

To introduce a covariant differentiation for other quantities than scalars, we define the parallel displacement of line-elements and the parallel displacement of quantities from one line-element to another parallel. By analogy with the Riemann space, it is natural to take as coefficients of a connection in the Finsler space the functions

$$\begin{cases} \Gamma^{*k}_{ij} = \Gamma^{*k}_{ji} = g^{kh}\Gamma^*_{ihj}, \\ \text{where} \\ \Gamma^*_{ihj} = \tfrac{1}{2}(g_{ih(j)} + g_{jh(i)} - g_{ij(h)}). \end{cases} \tag{2.8}$$

From (1.3) and (2.4) it follows that

$$\Gamma^{*k}_{ij} x'^i = \frac{\partial G^k}{\partial x'^j}, \qquad \Gamma^{*k}_{ij} x'^i x'^j = 2G^k. \tag{2.9}$$

The parallel displacement of the line-element (x, x') to the point $(x + dx)$ is now defined by

$$dl^i + l^h \Gamma^{*i}_{hj}\, dx^j = dl^i + \frac{1}{F} \frac{\partial G^i}{\partial x'^j} dx^j = 0, \tag{2.10}$$

and the parallel displacement of an arbitrary vector X^i from its line-element (x, x') to the line-element obtained from (x, x') by parallel displacement to the point $(x + dx)$ by

$$dX^i + X^h \Gamma^{*i}_{hj}\, dx^j = 0. \tag{2.11}$$

The left member of (2.11) is the invariant differential DX^i of the vector X^i. For any quantity, it is defined in a corresponding way, as is well known. The

coefficient of dx^j in the invariant differential of a quantity Φ is called the covariant derivative $\Phi_{|j}$ of the quantity. For instance, we have

$$(2.12)\qquad \begin{cases} X^i_{|j} = X^i_{(j)} + X^h \Gamma^{*i}_{hj}, \\ T_{ik|j} = T_{ik(j)} - T_{rk}\Gamma^{*r}_{ij} - T_{ir}\Gamma^{*r}_{kj}, \end{cases}$$

where $X^i_{(j)}$, $T_{ik(j)}$ are given by (2.6). (2.12) defines the covariant derivative also for sets of functions which behave like tensors under transformations of coordinates, but are not homogeneous of degree zero in x'. For a scalar, the differentiations $\cdots_{(i)}$ and $\cdots_{|i}$ are identical.

From the definition of the covariant derivative and from (2.7), (2.10)–(2.12), it follows immediately that

$$(2.13)\qquad \begin{cases} \text{(a)}\quad F_{|j} = 0, \qquad\qquad \text{(b)}\quad g_{ik|j} = 0, \\ \qquad \text{(c)}\quad l^i_{|j} = 0, \qquad l_{i|j} = 0. \end{cases}$$

Besides the covariant derivative we use in the following the derivative

$$(2.14)\qquad \Phi_{||j} = F\frac{\partial \Phi}{\partial x'^j}.$$

For a function Φ homogeneous of degree zero in x', we have

$$(2.15)\qquad \Phi_{||0} = 0.$$

Both differentiations $\cdots_{|j}$ and $\cdots_{||j}$ generate from a given quantity a new quantity of the same kind having a subscript more.

Finally, we state the *connection between* Γ^{*k}_{ij} *and* G^k. For that purpose we consider the symmetric tensor of the third order

$$(2.16)\qquad A_{ikj} = \tfrac{1}{2}g_{ik||j} = \tfrac{1}{4}F\frac{\partial^3(F^2)}{\partial x'^i\,\partial x'^k\,\partial x'^j}.$$

We have then

$$(2.17)\qquad A^{ik}_j = -\tfrac{1}{2}g^{ik}_{||j},$$

and with regard to the homogeneity of F

$$(2.18)\qquad A_{ik0} = A_{i0j} = A_{0kj} = 0.$$

If we write

$$(2.19)\qquad \frac{\partial^2 G^k}{\partial x'^i\,\partial x'^j} = G^k_{ij},$$

the connection sought is given by

$$(2.20)\qquad \Gamma^{*k}_{ij} = G^k_{ij} - A^k_{ij|0}.$$ [9]

[9] See E. Cartan, [6], VIII.

3. Permutation formulas. The differentiations $\cdots_{(j)}$ and $\cdots_{\|j}$ satisfy the permutation formulas

$$\Phi_{(i)(k)} - \Phi_{(k)(i)} = -\Phi\|_r R^r_{0ik}, \tag{3.1}$$

where

$$R^r_{0ik} = \frac{1}{F}\left(\frac{\partial^2 G^r}{\partial x'^i \partial x^k} - \frac{\partial^2 G^r}{\partial x'^k \partial x^i} - G^r_{is}\frac{\partial G^s}{\partial x'^k} + G^r_{ks}\frac{\partial G^s}{\partial x'^i}\right), \tag{3.2}$$

and

$$\Phi_{(i)\|k} - \Phi_{\|k(i)} = -\Phi\|_r(A^r_{ik|0} + \Gamma^{*r}_{ik}). \tag{3.3}$$

R^r_{0ik} is a tensor which results from Cartan's curvature tensor R^r_{hik} by inner multiplication with l^h.[10] (3.3) is obtained by means of (2.7) and (2.20).

4. Two-dimensional Finsler spaces. Normal vector and main scalar. In a two-dimensional Finsler space, besides the unit vector (l), the *normal vector*, i.e., the unit vector orthogonal to (l), is of importance. If h_i are its covariant, h^i its contravariant components, we have

$$\text{(4.1)}\quad \begin{aligned} &\text{(a)}\ g_{ik} = l_i l_k + h_i h_k, \\ &\text{(b)}\ \delta^k_i = l_i l^k + h_i h^k, \quad (\delta^k_i = 1 \text{ for } i = k, = 0 \text{ for } i \neq k). \end{aligned}$$

The connection between the vectors (l) and (h) is given with the aid of the ϵ-tensor the components of which are

$$\text{(4.2)}\quad \begin{cases} \epsilon_{11} = 0, & \epsilon_{12} = \sqrt{g}, \quad \epsilon_{21} = -\sqrt{g}, \quad \epsilon_{22} = 0, \\ \epsilon^{11} = 0, & \epsilon^{12} = \dfrac{1}{\sqrt{g}}, \quad \epsilon^{21} = -\dfrac{1}{\sqrt{g}}, \quad \epsilon^{22} = 0. \end{cases}$$

Here and later on, $\sqrt{\ }$ indicates the *positive* root. It is seen that

$$\epsilon_{ik} + \epsilon_{ki} = 0, \qquad \epsilon^{ik}\epsilon_{jk} = \delta^i_j, \qquad \epsilon^{ik}\epsilon_{ik} = 2. \tag{4.3}$$

Now we obtain, if the orientation of the normal vector is suitably chosen

$$\text{(4.4)}\quad \begin{cases} \text{(a)} & l^i = \epsilon^{ik} h_k, \qquad l_i = \epsilon_{ik} h^k, \\ \text{(b)} & h^i = -\epsilon^{ik} l_k, \qquad h_i = -\epsilon_{ik} l^k, \\ \text{(c)} & l^i h^k - l^k h^i = \epsilon^{ik}, \qquad l_i h_k - l_k h_i = \epsilon_{ik}. \end{cases}$$

Using (2.13, b, c), we have $\epsilon_{ik|j} = \epsilon^{ik}_{|j} = 0$ and therefore

$$h^i_{|j} = 0, \qquad h_{i|j} = 0. \tag{4.5}$$

The normal vector can be used for expressing the tensor A_{ik}, by a scalar. With regard to (2.18) we can put

$$A_{ikj} = \tfrac{1}{2} g_{ik}\|_j = I h_i h_k h_j. \tag{4.6}$$

---

[10] E. Cartan, [6], formulas (XIX), (XX), p. 36.

Hence we have in particular

$$(4.7) \qquad (\sqrt{g})_{\|j} = I\sqrt{g}\, h_j .$$

I is called the *main scalar* of the two-dimensional Finsler space.[11]

5. The derivative Φ_b. By means of the normal vector we define the derivative

$$(5.1) \qquad \Phi_b = \Phi_{(i)} h^i .$$

Then we have conversely

$$(5.2) \qquad \Phi_{(i)} = \Phi_s l_i + \Phi_b h_i .$$

For a function Φ positively homogeneous of degree zero in x', the derivative Φ_b corresponds to the difference of its values in the line-element (x, x') and in the line-element obtained by parallel displacement in transversal direction. Since from (2.13c), (4.5) it follows that

$$(5.3) \qquad l^i_{(j)} = -\frac{1}{F}\frac{\partial G^i}{\partial x'^j}, \qquad h^i_{(j)} = -h^r \Gamma^{*i}_{rj}$$

and

$$(5.4) \qquad l_{i(j)} = l_{j(i)} = l_r G^r_{ij}, \qquad h_{i(j)} = h_{j(i)} = h_r \Gamma^{*r}_{ij},$$

we get

$$(5.5) \qquad l^i_s = -2\frac{G^i}{F^2}, \qquad l^i_b = h^i_s = -\frac{1}{F}\frac{\partial G^i}{\partial x'^r} h^r, \qquad h^i_b = -\Gamma^{*i}_{rs} h^r h^s,$$

and

$$(5.6) \qquad \begin{cases} (l_i)_s = \dfrac{1}{F}\dfrac{\partial G^r}{\partial x'^i} l_r, & (l_i)_b = G^r_{is} l_r h^s, \\[2ex] (h_i)_s = \dfrac{1}{F}\dfrac{\partial G^r}{\partial x'^i} h_r, & (h_i)_b = \Gamma^{*r}_{is} h_r h^s. \end{cases}$$

The first equation (5.5) is taken from (2.3).

6. Landsberg's angle and the derivative Φ_ϑ. In a two-dimensional Finsler space, we introduce for functions Φ, which are *positively homogeneous of degree zero in* x', in addition to the derivatives Φ_s, Φ_b the derivative Φ_ϑ with respect to Landsberg's angle. Landsberg[12] defines the angle at the point (x) by the integral

$$(6.1) \qquad \vartheta = \int \epsilon_{ik} \frac{x'^i dx'^k}{F^2} = \int h_k\, dl^k$$

[11] L. Berwald, [1], p. 204. There $\mathfrak{J} = \frac{1}{2}I$ is called the main scalar.

[12] G. Landsberg, [12].

up to an arbitrary additive function of position. The angle between two line-elements $(x, \bar{x}')$ and $(x, \bar{\bar{x}}')$ at the point (x) is defined by the definite integral (6.1), taken from $\bar{x}'^2:\bar{x}'^1$ to $\bar{\bar{x}}'^2:\bar{\bar{x}}'^1$.

Because of (6.1) we have $\vartheta_{||i} = h_i$ and therefore, for a function Φ positively homogeneous of degree zero in x'

$$\Phi_{||i} = \Phi_\vartheta h_i, \qquad \Phi_\vartheta = \Phi_{||i}h^i. \tag{6.2}$$

From (4.7) and (4.2) it follows in particular that

$$\begin{cases} (g_{ik})_\vartheta = 2Ih_ih_k, \\ (\sqrt{g})_\vartheta = I\sqrt{g}, \qquad \sqrt{g} = \varphi(x)e^{\int I\,d\vartheta}, \end{cases} \tag{6.3}$$

where $\varphi(x)$ is an arbitrary function of position, and

$$(\epsilon_{ik})_\vartheta = I\epsilon_{ik}, \qquad (\epsilon^{ik})_\vartheta = -I\epsilon^{ik}. \tag{6.4}$$

Further we get from (1.8), (1.9), (4.4)

$$\begin{cases} l^i_\vartheta = h^i, \qquad h^i_\vartheta = -l^i - Ih^i, \\ (l_i)_\vartheta = h_i, \qquad (h_i)_\vartheta = -l_i + Ih_i. \end{cases} \tag{6.5}$$

7. Curvature of a two-dimensional Finsler space. Cartan's permutation formulas. On account of (2.13b), the parallel displacement (2.11) is euclidean (or metrical). Therefore the curvature tensor $R_{hjik} = g_{rj}R^r_{hik}$ is skew-symmetric in the two first indices also.[13] Hence we have $R^r_{0ik}l_r = 0$ and

$$R^r_{0ik} = Kh^r\epsilon_{ik}. \tag{7.1}$$

The scalar

$$K = \tfrac{1}{2}R^r_{0ik}h_r\epsilon^{ik} \tag{7.2}$$

is called the *curvature* of the Finsler space.[14]

From the permutation formulas (3.1), (3.2) we can now derive *Cartan's permutation formulas*,[15] which are fundamental for the two-dimensional Finsler space. Above all we have from (4.5) and (4.6)

$$A^r_{ik|0} = I_sh_ih^rh_k. \tag{7.3}$$

Now let Φ be a function positively homogeneous of degree zero in x'. From (3.1), (3.3) and (7.1), (7.3) (6.2) it follows that

$$\Phi_{(i)(k)} - \Phi_{(k)(i)} = -\Phi_\vartheta K\epsilon_{ik}, \tag{7.4}$$

$$\Phi_{(i)||k} - \Phi_{||k(i)} = -\Phi_\vartheta(I_sh_ih_k + h_r\Gamma^{*r}_{ik}). \tag{7.5}$$

[13] E. Cartan, [6], p. 36.

[14] K was introduced by A. L. Underhill, [16].

[15] E. Cartan, [5], p. 121; see also P. Funk, [11].

By going back to the definitions of Φ_s, Φ_b, Φ_ϑ, we obtain with regard to (5.5) and (6.5) *Cartan's permutation formulas*

$$(7.6)\qquad \boxed{\begin{aligned} \Phi_{b\vartheta} - \Phi_{\vartheta b} &= -\Phi_s - I\Phi_b - I_s\Phi_\vartheta, \\ \Phi_{\vartheta s} - \Phi_{s\vartheta} &= \qquad - \Phi_b \\ \Phi_{sb} - \Phi_{bs} &= \qquad\qquad - K\Phi_\vartheta. \end{aligned}}$$

From (7.6) and from Jacobi's identity

$$(7.7)\qquad \{(\Phi_{sb\vartheta} - \Phi_{bs\vartheta}) - (\Phi_{\vartheta sb} - \Phi_{\vartheta bs})\} + \text{cycl.} = 0,$$

where + cycl. stands for the terms which arise by cyclical permutation of s, b, ϑ from the term written, we obtain, by cancelling a factor Φ_ϑ, the "Bianchi" identity[16]

$$(7.8)\qquad \boxed{I_{ss} + IK + K_\vartheta = 0.}$$

8. The tensor FG^i_{jkr}. Affinely connected Finsler spaces. In the following we put, for the sake of simplicity,

$$(8.1)\qquad \frac{\partial^3 G^i}{\partial x'^j \partial x'^k \partial x'^r} = G^i_{jkr}, \qquad \frac{\partial^4 G^i}{\partial x'^j \partial x'^k \partial x'^r \partial x'^t} = G^i_{jkrt}$$

In consequence of the homogeneity of G^i we have $G^i_{0kr} = G^i_{j0r} = G^i_{jk0} = 0$. Therefore if we decompose the tensor FG^i_{jkr} in components according to the vectors (l) and (h), there will appear only terms with $l^ih_jh_kh_r$ and $h^ih_jh_kh_r$.

By carrying through this decomposition, we find first from (2.19), (2.20), (7.3), (6.2)

$$(8.2)\qquad FG^i_{jkr} = \Gamma^{*i}_{jk}{}_{||r} + (I_sh^ih_jh_k)_\vartheta h_r.$$

With regard to (6.5) we obtain

$$(8.3)\qquad (I_sh^ih_jh_k)_\vartheta = (I_{s\vartheta} + II_s)h^ih_jh_k - I_s(l^ih_jh_k + h^il_jh_k + h^ih_jl_k).$$

On the other hand, we have from (2.8), (2.17), (4.6)

$$(8.4)\qquad \Gamma^{*i}_{jk}{}_{||r} = g^{ih}\Gamma^*_{jhk||r} - 2Ih^ih_rh_m\Gamma^{*m}_{jk}.$$

From (2.8), (7.5) and (4.6), (5.4) we get for the first term on the right of (8.4)

$$(8.5)\qquad \begin{aligned} g^{ih}\Gamma^*_{jhk||r} = [(I_b - II_s)h^ih_jh_k - I_s(l^ih_jh_k \\ - h^il_jh_k - h^ih_jl_k)]h_r + 2Ih^ih_rh_m\Gamma^{*m}_{jk}. \end{aligned}$$

Hence we obtain finally

$$(8.6)\qquad \boxed{FG^i_{jkr} = [-2I_sl^i + (I_{s\vartheta} + I_b)h^i]h_jh_kh_r.}$$

[16] It appears first, expressed otherwise, in L. Berwald, [1], formula (78), p. 206.

In particular we have

$$FG^r_{jkr} = (I_{s\vartheta} + I_b)h_jh_k. \tag{8.7}$$

By operating on (8.7) with $\cdots_{\|m} = h_m \frac{\partial}{\partial\vartheta}$, it follows with regard to (1.9), (6.5) that

$$\begin{aligned} F^2G^r_{jkmr} &= [(I_{s\vartheta} + I_b)_\vartheta + 2I(I_{s\vartheta} + I_b)]h_jh_kh_m \\ &\quad -(I_{s\vartheta} + I_b)(l_jh_kh_m + h_jl_kh_m + h_jh_kl_m). \end{aligned} \tag{8.8}$$

A Finsler space is said to be *affinely connected*, if G^i are quadratic polynomials in x'. From (8.6) it is seen that the two-dimensional affinely connected Finsler spaces are characterized by

$$I_s = I_b = 0. \tag{8.9}$$

According to the third permutation formula (7.6) we have for these spaces either $K = 0$ or $I =$ const. From (3.2), (7.2) we see that, for an affinely connected Finsler space with $K = 0$, the G^i_{jk}, which are functions of position only, can all be transformed together to zero by a transformation of coordinates. Then it follows from (2.7) that F does not depend on x. The converse is evident. Hence the affinely connected Finsler spaces with $K = 0$ are identical with the *Minkowski spaces*. The two-dimensional affinely connected Finsler spaces with $I =$ const. were determined by the author. In §14 we come back to these spaces.[17]

II. Necessary and Sufficient Conditions for Two-dimensional Finsler Spaces With Rectilinear Extremals

9. Differential equations for the vectors (l) and (h) in a two-dimensional Finsler space with rectilinear extremals. The definition of a Finsler space with rectilinear extremals was given in the introduction. In the coordinate system mentioned there, the differential equations of the extremals are

$$x'^i x''^k - x'^k x''^i = 0, \qquad \left(x'^m = \frac{dx^m}{dt}, x''^m = \frac{d^2x^m}{dt^2}\right). \tag{9.1}$$

For a two-dimensional Finsler space there exists but one such equation. In this section, we seek conditions, invariant under transformation of coordinates, which are necessary and sufficient in order that a two-dimensional Finsler space may have rectilinear extremals.

First we suppose that the considered Finsler space has rectilinear extremals. Let x^i be coordinates for which the extremals are given by (9.1). It is evident that (9.1) is equivalent to

$$G^i = F^2ql^i, \tag{9.2}$$

[17] Cf. L. Berwald, [1], pp. 207 f., 212 f., 215 ff.; E. Cartan, [5], p. 134 f.

where q is a function of x, x', positively homogeneous of degree zero in x'. The The basis of our considerations is the system of equations (5.5), (6.5). On account of (9.2), it can be written as

$$(9.3) \qquad l_s^i = -2ql^i, \qquad l_b^i = -q_\vartheta l^i - qh^i, \qquad l_\vartheta^i = h^i;$$

$$(9.4) \qquad \begin{aligned} h_s^i = -q_\vartheta l^i - qh^i, \qquad h_b^i &= -(q + Iq_\vartheta + q_{\vartheta\vartheta})l^i - (2q_\vartheta - I_s)h^i, \\ h_\vartheta^i &= -l^i - Ih^i. \end{aligned}$$

In order to derive (9.3), (9.4), we can proceed from the first and the last equation (9.3). Then the second permutation formula (7.6) gives the second equation (9.3), which is identical with the first equation (9.4). By the same procedure we obtain the second equation (9.4) from the first and the third.

Let us now consider the conditions of integrability of (9.3), (9.4). Two of them are identically satisfied, in consequence of the derivation of (9.3), (9.4) just given. The condition of integrability, which follows from the permutation formula for $l_{b\vartheta}^i - l_{\vartheta b}^i$, is also identically satisfied. Each of the other permutation formulas gives two conditions of integrability, since the coefficients of l^i and h^i must be respectively equal in both members. Thus we obtain six conditions of integrability:

$$(9.5) \qquad q_{\vartheta s} = 2q_b + qq_\vartheta\,.$$

$$(9.6) \qquad q^2 + q_s = -K,$$

$$(9.7) \qquad q_{\vartheta b} = (q + Iq_\vartheta + q_{\vartheta\vartheta})_s - q(q + Iq_\vartheta + q_{\vartheta\vartheta}) + I_s q_\vartheta - K - q_\vartheta^2\,,$$

$$(9.8) \qquad qq_\vartheta - q_b + 2q_{\vartheta s} = I_{ss} + KI = -K_\vartheta\,,$$

$$(9.9) \qquad (q + Iq_\vartheta + q_{\vartheta\vartheta})_\vartheta + 2I(q + Iq_\vartheta + q_{\vartheta\vartheta}) + 2I_s = 0,$$

$$(9.10) \qquad 3(q + Iq_\vartheta + q_{\vartheta\vartheta}) = I_{s\vartheta} + I_b\,.$$

In (9.8) we made use of the identity (7.8).

10. Reduction of the integrability conditions. In the first place, we show that the conditions of integrability (9.5)–(9.8) can be substituted by the two equations

$$(10.1) \qquad K = -q^2 - q_s\,,$$

$$(10.2) \qquad K_\vartheta = -3(q_b + qq_\vartheta),$$

the second of which is found by substituting the value of $q_{\vartheta s}$ from (9.5) in (9.8).

Indeed, (9.5), (9.7), (9.8) follow from (10.1), (10.2). For, by substituting the value of K_ϑ from (10.1) in (10.2) and making use of the second formula (7.6), we obtain (9.5), and by substituting the value of $3(q_b + qq_\vartheta)$ from (9.5) in (10.2), we get (9.8). In order to derive (9.7) we differentiate (9.5) with respect to ϑ and apply the second permutation formula (7.6) for $\Phi = q_\vartheta$. Thus we find

$$q_{\vartheta\vartheta s} = 2q_{b\vartheta} - q_{\vartheta b} + q_\vartheta^2 + qq_{\vartheta\vartheta} = 2(q_{b\vartheta} - q_{\vartheta b}) + q_{\vartheta b} + q_\vartheta^2 + qq_{\vartheta\vartheta}\,.$$

Now apply the first permutation formula (7.6) to $q_{b\vartheta} - q_{\vartheta b}$ and form $(q + Iq_\vartheta + q_{\vartheta\vartheta})_s$. Then equation (9.7) results easily, in consequence of (9.5) and (10.1).

11. Necessary and sufficient conditions for rectilinear extremals. From §10 we see that the conditions of integrability (9.5)–(9.10) of the system (9.3), (9.4) of partial differential equations may be reduced to (10.1), (10.2), (9.10), (9.9). The first three of these equations can be written as

$$q_s = -K - q^2, \tag{11.1}$$

$$q_b = -\tfrac{1}{3}K_\vartheta - qq_\vartheta, \tag{11.2}$$

$$q_{\vartheta\vartheta} = \tfrac{1}{3}(I_{s\vartheta} + I_b) - Iq_\vartheta - q. \tag{11.3}$$

The equations (9.10), (9.9) are equivalent to (11.3) and

$$(I_{s\vartheta} + I_b)_\vartheta + 2I(I_{s\vartheta} + I_b) + 6I_s = 0. \tag{11.4}$$

(11.1)–(11.3) are a system of partial differential equations for the function q. In consequence of the permutation formulas (7.6), the conditions of integrability of this system are obtained by calculating the expressions

$$A = q_{sb} - q_{bs} + Kq_\vartheta,$$

$$B = q_{\vartheta s\vartheta} - q_{\vartheta\vartheta s} - q_{\vartheta b},$$

$$C = q_{\vartheta b\vartheta} - q_{\vartheta\vartheta b} + q_{\vartheta s} + Iq_{\vartheta b} + I_s q_{\vartheta\vartheta}$$

with the aid of (11.1)–(11.3) and by equating the results to zero. The calculation of A, B, C is simplified by the remark that $q_{\vartheta s}$ is given by (9.5), $q_{\vartheta b}$ by (9.7), if q_b, $q + Iq_\vartheta + q_{\vartheta\vartheta}$ are respectively replaced by their values (11.2), (9.10).

A simple calculation shows that $A = 0$ reduces to

$$K_{\vartheta s} - 3K_b = 0. \tag{11.5}$$

For B, we find first

$$B = -\tfrac{2}{3}[K_{\vartheta\vartheta} + (I_{s\vartheta} + I_b)_s + IK_\vartheta].$$

Since the permutation formulas (7.6) give

$$I_{s\vartheta s} + I_{bs} = I_{ss\vartheta} - I_{sb} + I_{bs} = I_{ss\vartheta} + KI_\vartheta,$$

we have further

$$B = -\tfrac{2}{3}(K_\vartheta + IK + I_{ss})_\vartheta.$$

Because of the Bianchi identity (7.8), B is therefore identically zero. For the calculation of C, the second permutation formula (7.6) gives

$$(I_{s\vartheta} + I_b)_{s\vartheta} - (I_{s\vartheta} + I_b)_b = (I_{s\vartheta} + I_b)_{\vartheta s}.$$

By means of this formula, we obtain

$$C = \tfrac{1}{3}[(I_{s\vartheta} + I_b)_\vartheta + 2I(I_{s\vartheta} + I_b)]_s$$
$$- \tfrac{1}{3}q[(I_{s\vartheta} + I_b)_\vartheta + 2I(I_{s\vartheta} + I_b) + 6I_s] - 2K_\vartheta - 2IK,$$

or, with regard to (7.8)

$$C = \tfrac{1}{3}[(I_{s\vartheta} + I_b)_\vartheta + 2I(I_{s\vartheta} + I_b) + 6I_s]_s$$
$$- \tfrac{1}{3}q[(I_{s\vartheta} + I_b)_\vartheta + 2I(I_{s\vartheta} + I_b) + 6I_s].$$

Consequently C is zero, if (11.4) holds. Hence we have:

If (11.4), (11.5) *are satisfied, the system* (11.1)–(11.3) *is completely integrable.*

It is evident that the equations (11.4), (11.5) are necessary in order that a two-dimensional Finsler space may have rectilinear extremals. They are also sufficient. For let (11.4), (11.5) hold. Then the system (11.1)–(11.3) is completely integrable. For a solution q of this system, all conditions of integrability (9.5)–(9.10) of the system (9.3), (9.4) are satisfied. But the first equation (9.3) or

$$x^i_{ss} + 2qx^i_s = 0 \tag{11.6}$$

states that the considered two-dimensional Finsler space has rectilinear extremals. Therefore we have:

A necessary and sufficient condition in order that the extremals of a two-dimensional Finsler space be rectilinear is that the equations

$$\boxed{(I_{s\vartheta} + I_b)_\vartheta + 2I(I_{s\vartheta} + I_b) + 6I_s = 0} \tag{11.4}$$

and

$$\boxed{K_{\vartheta s} - 3K_b = 0} \tag{11.5}$$

be satisfied.[18]

12. Geometrical meaning of (11.4). In the foregoing paragraphs, we established the conditions (11.4), (11.5) in a purely analytical way which does not show the independent geometrical meaning of each of these conditions. We give therefore still another deduction, which does not possess this disadvantage.

For that purpose we insert some general considerations, restricting ourselves to the case of two dimensions, for the sake of brevity. Let a two-dimensional manifold with coordinates x^i be given and a differential equation

$$x'^h(x''^i + 2G^i(x, x')) - x'^i(x''^h + 2G^h(x, x')) = 0, \qquad \left(x'^m = \frac{dx^m}{dt},\ x''^m = \frac{d^2x^m}{dt^2}\right), \tag{12.1}$$

[18] Cf. P. Funk, [11].

where the functions $G^i(x, x')$ are analytic and positively homogeneous of the second degree in x'.[19] (12.1) defines a system of general *paths* $x^i = x^i(t)$ in the manifold. This system remains unaltered under the transformation

$$\bar{G}^i(x, x') = G^i(x, x') + p(x, x')x'^i, \tag{12.2}$$

where p is analytic and positively homogeneous of the first degree in x'.[19] We call (12.2) a *projective change* of the functions G^i. A transformation of coordinates $\overset{*}{x}{}^i = \overset{*}{x}{}^i(x)$ transforms $2G^i$ in

$$2\overset{*}{G}{}^h = \left(2G^i + \frac{\partial^2 x^i}{\partial \overset{*}{x}{}^j \partial \overset{*}{x}{}^k} \overset{*}{x}{}'^j \overset{*}{x}{}'^k\right)\frac{\partial \overset{*}{x}{}^h}{\partial x^i}. \tag{12.3}$$

The second and the higher derivatives

$$\frac{\partial^2 G^i}{\partial x'^j \partial x'^k} = G^i_{jk}; \qquad \frac{\partial^3 G^i}{\partial x'^j \partial x'^k \partial x'^s} = G^i_{jks}, \qquad \frac{\partial^4 G^i}{\partial x'^j \partial x'^k \partial x'^s \partial x'^m} = G^i_{jksm} \tag{12.4}$$

therefore transform respectively like the coefficients of an affine connection (depending upon the line-element), and like tensors. To the projective change (12.2) of G^i corresponds respectively the projective change

$$(12.5)\quad \begin{cases} \text{(a)}\ \bar{G}^i_{jk} = G^i_{jk} + \dfrac{\partial p}{\partial x'^j}\delta^i_k + \dfrac{\partial p}{\partial x'^k}\delta^i_j + \dfrac{\partial^2 p}{\partial x'^j \partial x'^k}x'^i, \\[2ex] \text{(b)}\ \bar{G}^i_{jks} = G^i_{jks} + \dfrac{\partial^2 p}{\partial x'^j \partial x'^k}\delta^i_s + \dfrac{\partial^2 p}{\partial x'^k \partial x'^s}\delta^i_j \\[2ex] \qquad\qquad + \dfrac{\partial^2 p}{\partial x'^s \partial x'^j}\delta^i_k + \dfrac{\partial^3 p}{\partial x'^j \partial x'^k \partial x'^s}x'^i \end{cases}$$

of the affine connection G^i_{jk}, and of the tensor G^i_{jks}.

It is evident that the extremals of a Finsler space form a system of paths, G^i being defined by (2.4) in this case. Now the meaning of the condition (11.4) can be expressed as follows:

A necessary and sufficient condition in order that in a two-dimensional Finsler space the functions G^i, defined by (2.4), *can be projectively changed in such a way that the transformed functions $\bar{G}^i$ be quadratic polynomials in x', is that equation* (11.4) holds.

PROOF: In the first place, we consider a two-dimensional space with a system of paths (12.1). Let $\bar{G}^i$ be quadratic polynomials in x. Then the right member of (12.5b) is zero. Now we have from (12.2), in consequence of the homogeneity of p

$$p = \frac{1}{3}\left(\frac{\partial \bar{G}^r}{\partial x'^r} - \frac{\partial G^r}{\partial x'^r}\right). \tag{12.6}$$

---

[19] We suppose here positive homogeneity with regard to the following applications to Finsler spaces. In general, mere homogeneity is supposed.

By substituting this value in (12.5b), we find as a necessary condition for $\bar{G}^i = \bar{G}^i_{kj}(x)x'^k x'^j$

$$G^i_{jks} - \tfrac{1}{3}(G^r_{r\,jk}\delta^i_s + G^r_{rks}\delta^i_j + G^r_{rsj}\delta^i_k) - \tfrac{1}{3}G^r_{r\,jks}x'^i = 0. \tag{12.7}$$

Conversely, if (12.7) holds, we put

$$p = \frac{1}{3}\left(2\bar{G}^r_{kr}(x)x'^k - \frac{\partial G^r}{\partial x'^r}\right), \tag{12.8}$$

where $\bar{G}^i_{kj} = \bar{G}^i_{jk}$ are arbitrary functions of x alone. Then from (12.7) and (12.5b) we have $\bar{G}^i_{jks} = 0$. Therefore (12.7) is sufficient too for $\bar{G}^i = \bar{G}^i_{kj}(x)x'^k x'^j$.[20]

Now consider in particular a two-dimensional Finsler space, and let G^i be given by (2.4). Then, by multiplying (12.7) by $-3F$ and taking account of (8.6)-(8.8), (4.1b), we obtain (11.4) multiplied by $l^i h_j h_k h_s$.

We can give the theorem just proved the form:

A necessary and sufficient condition in order that the extremals of a two-dimensional Finsler space form a quasigeodesic system of curves is that (11.4) *be satisfied.*[21]

It is easy to prove this form of the theorem directly. We have to suppose that for the extremals $x'^i x''^k - x'^k x''^i$ is a homogeneous cubic polynomial in x', or that

$$x'^i G^k - x'^k G^i = \tfrac{1}{6}\Phi^{ik}_{jhr}(x)x'^j x'^h x'^r. \tag{12.9}$$

If we differentiate (12.9) with respect to x'^j, x'^p, x'^r, x'^s, we obtain an equation which reduces to (11.4), multiplied by $\dfrac{1}{\sqrt{g}}\, h_j h_p h_r h_s$, as can be seen from §8 and §4.

13. Geometrical meaning of (11.5). Before considering (11.5), we observe that the system of functions

$$K^i_{jk} = \frac{\partial^2 G^i}{\partial x'^j \partial x^k} - \frac{\partial^2 G^i}{\partial x'^k \partial x^j} - G^i_{jr}\frac{\partial G^r}{\partial x'^k} + G^i_{kr}\frac{\partial G^r}{\partial x'^j}, \tag{13.1}$$

where G^i are the functions which enter in (12.1), behaves by a transformation of coordinates like a tensor, but is positively homogeneous of the *first* degree in x'. We call K^i_{jk} the *fundamental curvature tensor* of the connection G^i_{jk}. For a two-dimensional Finsler space, where G^i are defined by (2.4), the fundamental curvature tensor is connected with the curvature K by

$$K^i_{jk} = FKh^i\epsilon_{jk}, \tag{13.2}$$

as (3.2) and (7.1) show.

[20] See J. Douglas, [7], p. 157 f., for $n \geqq 2$ dimensions.

[21] Cf. P. Funk, [11]. For the name "quasigeodesic system of curves" see W. Blaschke and G. Bol, [4], p. 245.

Now the geometrical meaning of (11.5) is expressed by the theorem:
In order that there may exist a projective change of the G's defined by (2.4) *which transforms the fundamental tensor of curvature of a two-dimensional Finsler space to zero, it is necessary and sufficient that equation* (11.5) *be satisfied.*

To prove this theorem, we first establish a necessary and sufficient condition that in a two-dimensional manifold, bearing a system (12.1) of paths, the fundamental curvature tensor may be transformed to zero, by a projective change of G^i. Then we suppose that the considered manifold is a two-dimensional Finsler space, and that (12.1) is the system of its extremals (2.2), and we show that the necessary and sufficient condition we found reduces to (11.5) if we introduce the curvature K and its derivatives with respect to s, b, ϑ.

Let (12.1) be the system of paths of a two-dimensional manifold. We denote the covariant derivative with respect to the G^i_{jk} by a semicolon; for instance:

$$T_{j;k} = \frac{\partial T_j}{\partial x^k} - \frac{\partial T_j}{\partial x'^r}\frac{\partial G^r}{\partial x'^k} - T_r G^r_{jk}. \tag{13.3}$$

Let $\bar{K}^i_{jk}$ be the fundamental curvature tensor of the connection $\bar{G}^i_{jk}$ obtained from the connection G^i_{jk} by the projective change (12.2). Then we have

$$\begin{aligned}\bar{K}^i_{jk} = K^i_{jk} &- \left(p_{;j} - \frac{1}{2}\frac{\partial(p^2)}{\partial x'^j}\right)\delta^i_k \\ &+ \left(p_{;k} - \frac{1}{2}\frac{\partial(p^2)}{\partial x'^k}\right)\delta^i_j - \left(\frac{\partial p_{;j}}{\partial x'^k} - \frac{\partial p_{;k}}{\partial x'^j}\right)x'^i.\end{aligned} \tag{13.4}$$

Hence the postulate $\bar{K}^i_{jk} = 0$ gives

$$K^i_{jk} = \left(p_{;j} - \frac{1}{2}\frac{\partial(p^2)}{\partial x'^j}\right)\delta^i_k - \left(p_{;k} - \frac{1}{2}\frac{\partial(p^2)}{\partial x'^k}\right)\delta^i_j + \left(\frac{\partial p_{;j}}{\partial x'^k} - \frac{\partial p_{;k}}{\partial x'^j}\right)x'^i. \tag{13.5}$$

Suppose now that (13.5) is satisfied. If we put

$$K^r_{jr} = K_j, \qquad \frac{\partial K_r}{\partial x'^j} = K_{jr}, \tag{13.6}$$

we obtain by differentiation and contraction the system of differential equations

$$p_{;j} - \frac{1}{2}\frac{\partial(p^2)}{\partial x'^j} = \tfrac{2}{3}K_j + \tfrac{1}{3}K_{jr}x'^r. \tag{13.7}$$

for the function p. The condition of integrability of (13.7) is

$$\begin{aligned}p_{;j;k} - p_{;k;j} - \frac{1}{2}\left(\frac{\partial(p^2)}{\partial x'^j}\right)_{;k} + \frac{1}{2}\left(\frac{\partial(p^2)}{\partial x'^k}\right)_{;j} \\ = \tfrac{2}{3}(K_{j;k} - K_{k;j}) + \tfrac{1}{3}[(K_{jr}x'^r)_{;k} - (K_{kr}x'^r)_{;j}].\end{aligned} \tag{13.8}$$

In consequence of the permutation formulas

$$\left\{\begin{aligned}&\Phi_{;j;k}-\Phi_{;k;j}=-K^i_{jk}\frac{\partial\Phi}{\partial x'^i},\\&\left(\frac{\partial\Phi}{\partial x'^j}\right)_{;k}-\frac{\partial\Phi_{;k}}{\partial x'^j}=0,\end{aligned}\right.\tag{13.9}$$

and of the homogeneity of p, we find for the left member of (13.8) the value

$$-\frac{\partial p}{\partial x'^i}\left\{K^i_{jk}-\left(p_{;j}-\frac{1}{2}\frac{\partial(p^2)}{\partial x'^j}\right)\delta^i_k+\left(p_{;k}-\frac{1}{2}\frac{\partial(p^2)}{\partial x'^k}\right)\delta^i_j-\left(\frac{\partial p_{;j}}{\partial x'^k}-\frac{\partial p_{;k}}{\partial x'^j}\right)x'^i\right\},$$

which is zero, because of (13.5). Hence the condition of integrability of (13.7) becomes finally

$$\tfrac{2}{3}(K_{j;k}-K_{k;j})+\tfrac{1}{3}[(K_{jr}x'^r)_{;k}-(K_{kr}x'^r)_{;j}]=0.\tag{13.10}$$

Conversely, let (13.10) be satisfied. Then the condition of integrability of (13.7) leads back to (13.5). Hence (13.10) is a necessary and sufficient condition in order that the fundamental curvature tensor of a two-dimensional manifold with a system (12.1) of paths can be transformed to zero by a projective change of G^i.[22]

If the curves (12.1) are the extremals of a two-dimensional Finsler space, we have with regard to (2.6), (13.3) and to the symmetry $G^i_{jk}=G^i_{kj}$

$$T_{j;k}-T_{k;j}=T_{j(k)}-T_{k(j)}.\tag{13.11}$$

Further it follows from (13.2), (4.4), (6.2), (6.5), (4.1) that

$$K_j=FKl_j,\qquad K_{jr}=Kg_{rj}+K_\vartheta l_r h_j,\qquad K_{jr}x'^r=F(Kl_j+K_\vartheta h_j).\tag{13.12}$$

If we introduce, by means of (5.2), the derivatives with respect to s and b instead of the derivative $\cdots_{(i)}$, we obtain in consequence of (2.7), (5.4), (4.4) and of the symmetry $G^i_{jk}=G^i_{kj}$

$$\left\{\begin{aligned}&K_{j(k)}-K_{k(j)}=FK_b(l_jh_k-l_kh_j)=FK_b\epsilon_{jk},\\&(K_{rj}x'^r)_{(k)}-(K_{rk}x'^r)_{(j)}=F(K_b-K_{\vartheta s})\epsilon_{jk}.\end{aligned}\right.\tag{13.13}$$

By substituting these values in (13.10) and by dropping the factor $-\frac{1}{3}F\epsilon_{jk}$, we get (11.5).

From the theorems we proved in §§12, 13, we can also easily see that the conditions (11.4) and (11.5) are necessary and sufficient in order that the

[22] In consequence of these considerations and the corresponding ones in §12, we have incidentally the theorem:

The conditions (12.7), (13.10) *are necessary and sufficient in order that a given system of paths* (12.1) *of a two-dimensional manifold be equivalent by point transformation to the straight lines of a flat projective space.*

Cf. also the end of §13. For $n>2$ dimensions, J. Douglas, [7], p. 162 established the corresponding theorem, which is not however applicable when $n=2$.

extremals of a two-dimensional Finsler space be rectilinear. For let the extremals be rectilinear. Then, in a suitable coordinate system, (9.2) holds. Hence we obtain from (12.2) $\bar{G}^i = 0$, by putting $p = -qF$. Therefore (12.7) and (13.10) or, what amounts to the same thing, (11.4) and (11.5) are satisfied.

Conversely, if (11.4) and (11.5) are satisfied, the coefficients $\bar{G}^i_{jk}$ of the projectively transformed affine connection (2.19) are functions of position only. It is easily seen that the curvature tensor of the connection $\bar{G}^i_{jk}$ is $\dfrac{\partial \bar{K}^i_{jk}}{\partial x'^r}$. Since (11.5) holds, this tensor is zero. Therefore the $\bar{G}^i_{jk}$ can altogether be transformed to zero by a suitable transformation of coordinates $x^i = x^i(\overset{*}{x})$. Then

$$\overset{*}{G}{}^i_{jk} = 0, \qquad 2\overset{*}{G}{}^i = \overset{*}{G}{}^i_{jk} x'^j x'^k = 0. \tag{13.14}$$

Hence, in the new coordinate system, the differential equation of the extremals has the form

$$\overset{*}{x}{}'^i \overset{*}{x}{}''^k - \overset{*}{x}{}'^k \overset{*}{x}{}''^i = 0, \tag{13.15}$$

that is, the extremals are rectilinear.

III. The Two-dimensional Finsler Spaces with Rectilinear Extremals the Main Scalar of Which is a Function of Position Only

14. The two-dimensional Finsler spaces with constant main scalar. We shall now determine all two-dimensional Finsler spaces with rectilinear extremals the main scalar of which is a function of position only. In order to avoid interruptions later on, we first reproduce briefly the manner in which *all* Finsler spaces with I = const. can be determined,[23] and we add a simple remark on these spaces.

The Finsler spaces with I = const. belong to the class of Finsler spaces for which $\dfrac{F^2}{\sqrt{g}}$ is a quadratic polynomial in x'. In fact, if we differentiate $\dfrac{F^2}{\sqrt{g}}$ three times with respect to x' and if we use (1.9), (2.14), (6.2)–6.5), we get successively

$$\frac{\partial}{\partial x'^i}\frac{F^2}{\sqrt{g}} = \frac{F}{\sqrt{g}}(2l_i - Ih_i), \tag{14.1}$$

$$\frac{\partial^2}{\partial x'^i \partial x'^k}\left(\frac{F^2}{\sqrt{g}}\right) = \frac{1}{\sqrt{g}}[2l_il_k - I(l_ih_k + h_il_k) + (2 - I_\vartheta)h_ih_k], \tag{14.2}$$

$$\frac{\partial^3}{\partial x'^i \partial x'^k \partial x'^j}\left(\frac{F^2}{\sqrt{g}}\right) = -\frac{1}{F\sqrt{g}}(I_{\vartheta\vartheta} + II_\vartheta)h_ih_kh_j. \tag{14.3}$$

The two-dimensional Finsler spaces, for which

$$\frac{F^2}{\sqrt{g}} = \mathfrak{g}_{ik}(x)x'^ix'^k, \qquad (\mathfrak{g}_{ik} = \mathfrak{g}_{ki}), \tag{14.4}$$

---

[23] L. Berwald, [1], p. 215 ff.

are therefore characterized by $I_{\vartheta\vartheta} + II_\vartheta = 0$. With regard to (14.2) and (4.1a), we have from (14.4)

$$\mathfrak{g}_{ik}(x) = \frac{1}{\sqrt{g}}\{g_{ik} - \tfrac{1}{2}I(l_ih_k + l_kh_i) - \tfrac{1}{2}i_\vartheta h_ih_k\}, \tag{14.5}$$

and from (14.5)

$$\mathfrak{g} = 1 - \tfrac{1}{4}I^2 - \tfrac{1}{2}I_\vartheta\,, \tag{14.6}$$

where $\mathfrak{g} = \det.\ (\mathfrak{g}_{ik})$.

Suppose now in particular $I = \text{const.}$ Then we have from (6.3)

$$\sqrt{g} = e^{I\vartheta} \tag{14.7}$$

since we can choose in (6.3) $\varphi(x) = 1$ without loss of generality for the following considerations. Therefore it follows from (14.4) that

$$F = \sqrt{\mathfrak{g}_{ik}x'^ix'^k}\, e^{\frac{1}{2}I\vartheta}. \tag{14.8}$$

Now it remains to determine ϑ as function of x'. For that purpose we distinguish the three cases

$$\mathfrak{g} = 1 - \tfrac{1}{4}I^2 \gtreqless 0, \tag{14.9}$$

and understand by $\alpha_ix'^i$, $\beta_ix'^i$ two linearly independent Pfaffians so that $\Delta = \alpha_1\beta_2 - \alpha_2\beta_1 > 0$. We put in the case

(a) $I^2 < 4$:

$$\mathfrak{g}_{ik}x'^ix'^k = \frac{\sqrt{\mathfrak{g}}}{\Delta}[(\alpha_ix'^i)^2 + (\beta_ix'^i)^2]; \tag{14.10a}$$

(b) $I^2 = 4$:

$$\mathfrak{g}_{ik}x'^ix'^k = \frac{1}{\Delta}(\alpha_ix'^i)^2; \tag{14.10b}$$

(c) $I^2 > 4$:

$$\mathfrak{g}_{ik}x'^ix'^k = \frac{2\sqrt{-\mathfrak{g}}}{\Delta}(\alpha_ix'^i)(\beta_kx'^k). \tag{14.10c}$$

Because of (6.1) and (14.1) we have

$$\vartheta = \int\frac{x'^1dx'^2 - x'^2dx'^1}{\mathfrak{g}_{ik}(x)x'^ix'^k}. \tag{14.11}$$

By integrating, we get respectively in the cases (a), (b), (c):

$$\vartheta = \frac{1}{\sqrt{\mathfrak{g}}}\operatorname{arctg}\frac{\beta_ix'^i}{\alpha_ix'^i} + \psi(x), \tag{14.12a}$$

$$\vartheta = \frac{\beta_ix'^i}{\alpha_ix'^i} + \psi(x), \tag{14.12b}$$

$$\vartheta = \frac{1}{2\sqrt{-g}} \log \frac{\beta_i x'^i}{\alpha_i x'^i} + \psi(x), \tag{14.12c}$$

where $\psi(x)$ is an arbitrary function of position. By substituting these values in (14.8), we obtain finally, by means of (14.9)–(14.12), the following types of *two-dimensional Finsler spaces with I = const.*

$$(14.13a) \quad I^2 < 4 \quad : \quad F = ((\alpha_i x'^i)^2 + (\beta_i x'^i)^2)^{\frac{1}{2}} e^{\frac{I}{\sqrt{(4-I^2)}} \operatorname{arctg} \frac{\beta_i x'^i}{\alpha_i x'^i}},$$

$$(14.13b) \quad I^2 = 4 \quad : \quad F = (\alpha_i x'^i) e^{\frac{1}{2} I \frac{\beta_i x'^i}{\alpha_i x'^i}},$$

$$(14.13c) \quad I^2 > 4 \quad : \quad F = (\alpha_i x'^i)^{\frac{1}{2}\left(1 - \frac{I}{\sqrt{(I^2-4)}}\right)} (\beta_i x'^i)^{\frac{1}{2}\left(1 + \frac{I}{\sqrt{(I^2-4)}}\right)}.$$

In consequence of (14.10) and (14.12) a multiplicative positive function of position would enter in (14.13). We have taken it in in α_i, β_i.

For the Finsler spaces with I = const. we have the following theorem:

If the curvature of a two-dimensional Finsler space with constant $I \neq 0$ is a function of position only, the space is a Minkowski space.

Indeed, because of I = const. the Finsler space is affinely connected (§8). With regard to I = const., $I \neq 0$, $K_\vartheta = 0$, the Bianchi identity (7.8) gives $K = 0$. Hence the space is a Minkowski space (§8).

15. Theorems on two-dimensional Finsler space with rectilinear extremals. In the first place, we establish some theorems which hold for Finsler spaces with rectilinear extremals.

THEOREM 1. *When for a two-dimensional Finsler space with a quasigeodesic system of extremals the main scalar is a function of position only, it is constant.*

PROOF: From the hypothesis

$$I_\vartheta = 0 \tag{15.1}$$

it follows, with the aid of the first two permutation formulas (7.6), that

$$I_{s\vartheta} = I_b, \qquad I_{b\vartheta} = -I_s - II_b. \tag{15.2}$$

By substituting these values in (11.4), we have

$$II_b + 2I_s = 0. \tag{15.3}$$

If we differentiate (15.3) with respect to ϑ and take note of (15.1), (15.2), we obtain

$$(I^2 - 2)I_b + II_s = 0. \tag{15.4}$$

The determinant of the equations (15.3), (15.4) for I_b, I_s is $4 - I^2$. Therefore we have either $I^2 = 4$ or $I_b = I_s = 0$, that is, in each case, I = const.

THEOREM 2. *A two-dimensional Finsler space with rectilinear extremals for which I^2 is a function of position only and $\neq \frac{9}{2}$, has constant curvature.*

PROOF: *With regard to theorem 1 we have by hypothesis*

$$(15.5)\qquad (a)\ I = \text{const.},\qquad (b)\ K_{\vartheta s} = 3K_b .$$

Because of (15.5a) the Bianchi identity (7.8) gives

$$(15.6)\qquad K_\vartheta = -IK.$$

In consequence of I = const., it follows from (15.6) that

$$(15.7)\qquad (a)\ K_{\vartheta s} = -IK_s ,\qquad (b)\ K_{\vartheta b} = -IK_b .$$

By substituting the value of $K_{\vartheta s}$ from (15.5b) in (15.7a), we find

$$(15.8)\qquad IK_s + 3K_b = 0.$$

If we differentiate (15.8) with respect to ϑ, and make use of (15.5), (15.7) and of the first two permutation formulas (7.6), we obtain

$$(15.9)\qquad 3K_s + 2IK_b = 0.$$

The determinant of the equations (15.8), (15.9) for K_b , K_s is $2I^2 - 9$. Accordingly we get from (15.8), (15.9) $K_b = K_s = 0$, when $I^2 \neq \frac{9}{2}$. Finally, the last permutation formula (7.6) gives $KK_\vartheta = 0$. Hence K = const.

As a corollary of theorem 2 and the theorem of §14, we have

THEOREM 3. *A two-dimensional Finsler space with rectilinear extremals for which I^2 is a function of position only and $\neq 0$, $\frac{9}{2}$ is a Minkowski space.*

For such a space is $K = 0$.

16. The exceptional case $I^2 = \frac{9}{2}$. We shall now determine the two-dimensional Finsler spaces with rectilinear extremals for which $I^2 = \frac{9}{2}$. Let us begin with some general remarks.

We write in the following

$$(16.1)\qquad x^1 = x,\qquad x^2 = y,$$

and we denote partial differentiation by subscripts. It is known that the differential equation of the extremals of a two-dimensional Finsler space can be written as follows

$$(16.2)\qquad F_{xy'} - F_{yx'} + F_1(x'y'' - y'x'') = 0,$$

where

$$(16.3)\qquad F_1 = \frac{F_{x'x'}}{(y')^2} = -\frac{F_{x'y'}}{x'y'} = \frac{F_{y'y'}}{(x')^2}.$$

From (16.3) we have

$$(16.4)\qquad F_{xy'} - F_{yx'} = 0$$

as a necessary and sufficient condition for rectilinear extremals. By differentiating (16.4) with respect to x' or y', we obtain, in consequence of (16.3), the

necessary condition

$$\frac{\partial F_1}{\partial x} x' + \frac{\partial F_1}{\partial y} y' = 0. \tag{16.5}$$

Now let us consider a two-dimensional Finsler space with $I^2 = \frac{9}{2}$. With regard to (14.13c) and (16.1), we have for such a space

$$F = \frac{(\alpha x' + \beta y')^2}{\gamma x' + \delta y'}, \qquad (\alpha\delta - \beta\gamma \neq 0). \tag{16.6}$$

We multiply the numerator and the denominator by $(\alpha\delta - \beta\gamma)^{-\frac{1}{2}}$ and put

$$\begin{aligned} A &= \frac{\alpha}{(\alpha\delta - \beta\gamma)^{\frac{1}{2}}}, & B &= \frac{\beta}{(\alpha\delta - \beta\gamma)^{\frac{1}{2}}}, \\ C &= \frac{\gamma}{(\alpha\delta - \beta\gamma)^{\frac{1}{2}}}, & D &= \frac{\delta}{(\alpha\delta - \beta\gamma)^{\frac{1}{2}}}. \end{aligned} \tag{16.7}$$

Then we obtain

$$F = \frac{(Ax' + By')^2}{Cx' + Dy'}, \tag{16.8}$$

where

$$AD - BC = 1. \tag{16.9}$$

Without loss of generality, we can suppose $D \neq 0$. For, if $D = 0$, $C \neq 0$, we arrive at $D \neq 0$ by permuting x and y. Then we have

$$A = \frac{1 + BC}{D} \tag{16.10}$$

and

$$F = \frac{1}{D^2}\left\{\frac{(x')^2}{Cx' + Dy'} + 2Bx' + B^2(Cx' + Dy')\right\}. \tag{16.11}$$

We start with the form (16.11) of the fundamental function F and firstly make use of the necessary condition (16.5). For the function (16.11) we have

$$F_1 = \frac{2}{(Cx' + Dy')^3}. \tag{16.12}$$

Consequently condition (16.5) runs as follows

$$\frac{\partial C}{\partial x}(x')^2 + \left(\frac{\partial C}{\partial y} + \frac{\partial D}{\partial x}\right)x'y' + \frac{\partial D}{\partial y}(y')^2 = 0. \tag{16.13}$$

Since C, D are functions of position, (16.13) drops into the system of differential equations

$$\frac{\partial C}{\partial x} = \frac{\partial C}{\partial y} + \frac{\partial D}{\partial x} = \frac{\partial D}{\partial y} = 0, \tag{16.14}$$

which gives, by integration,

(16.15) $$C = -ay + c, \qquad D = ax + d \qquad (a, c, d \text{ constants}).$$

Now we distinguish two cases:
(a) $a = 0$.

(16.15a) $$C = c, \qquad D = d.$$

(b) $a \neq 0$.

(16.15b) $$C = -a(y - y_0), \qquad D = a(x - x_0).$$

In case (a) we make the transformation of coordinates

(16.16a) $$\bar{x} = x, \qquad \bar{y} = d^2(cx + dy)$$

and put

(16.17a) $$\frac{1}{d^2} B(x, y) = \frac{1}{d^2} B\left(\bar{x}, \frac{\bar{y}}{d^3} - \frac{c}{d}\bar{x}\right) = z(\bar{x}, \bar{y}).$$

By substituting in (16.11) and dropping the bars, we obtain

(16.18) $$\boxed{F = \frac{(x')^2}{y'} + 2zx' + z^2y' = \frac{(x' + zy')^2}{y'}.}$$

In case (b) the transformation of coordinates

(16.16b) $$\bar{x} = \frac{1}{x - x_0}, \qquad \bar{y} = a^3 \frac{y - y_0}{x - x_0}$$

gives

$$C = -\frac{1}{a^2}\frac{\bar{y}}{\bar{x}}, \qquad D = a\frac{1}{\bar{x}}, \qquad x' = -\frac{\bar{x}'}{\bar{x}^2}, \qquad Cx' + Dy' = \frac{1}{a^2}\frac{\bar{y}'}{\bar{x}^2}.$$

Further we put

(16.17b) $$-\frac{1}{a^2} B(x, y) = -\frac{1}{a^2} B\left(\frac{1}{\bar{x}} + x_0, \frac{1}{a^3}\frac{\bar{y}}{\bar{x}} + y_0\right) = z(\bar{x}, \bar{y}).$$

By substituting these values in (16.11) and dropping the bars, we recover (16.18).

The fundamental function (16.18) satisfies the necessary condition (16.5), but not yet the necessary and sufficient condition (16.4). This condition gives for z the differential equation

(16.19) $$\boxed{z\frac{\partial z}{\partial x} - \frac{\partial z}{\partial y} = 0.}$$

(16.19) has the obvious solution $z = \text{const.}$ which gives a Minkowski space. If z is not constant, integration of (16.19) leads to

(16.20) $$\boxed{x + yz = \Psi(z).}$$

where Ψ is an arbitrary function of z (which we suppose to be analytic).

Hence we have:

The fundamental function of a two-dimensional Finsler space with rectilinear extremals and $I^2 = \frac{9}{2}$ *which is no Minkowski space, can be brought by a suitable transformation of coordinates, to the form* (16.18), *where* z *is a non constant solution of* (16.19). *These solutions are given by* (16.20), *where* $\Psi(z)$ *is an arbitrary function of* z.

In §17 we shall see that non constant solutions of (16.19) exist which nevertheless give a Minkowski space.

17. Space curvature in the exceptional case. We calculate now the curvature K of a Finsler space with rectilinear extremals and $I^2 = \frac{9}{2}$, using (11.1). The function

$$q = \frac{1}{F^2} l_i G^i \tag{17.1}$$

introduced by (9.2), is connected with F by

$$q = \frac{1}{2F^2} \frac{\partial F}{\partial x^i} x'^i \tag{17.2}$$

as is seen from

$$\frac{dF}{ds} = \frac{\partial F}{\partial x^i} \frac{x'^i}{F} - 2 \frac{1}{F} l_i G^i = 0. \tag{17.3}$$

Further we have from (9.2)

$$q_s = \frac{1}{F} \frac{\partial q}{\partial x^i} x'^i, \tag{17.4}$$

if we note the homogeneity of q in the x'.

Now let F be given by (16.18), (16.19). Then we substitute, with the aid of (16.19), the partial derivatives $\frac{\partial z}{\partial x}, \frac{\partial^2 z}{\partial x^2}$ for the derivatives $\frac{\partial z}{\partial y} \frac{\partial^2 z}{\partial y \partial x}$:

$$\frac{\partial z}{\partial y} = z \frac{\partial z}{\partial x}, \qquad \frac{\partial^2 z}{\partial y \partial x} = z \frac{\partial^2 z}{\partial x^2} + \left(\frac{\partial z}{\partial x}\right)^2. \tag{17.5}$$

Moreover it follows from (16.20), by partial differentiation with respect to x, that

$$\frac{\partial z}{\partial x} = \frac{1}{\Psi'(z) - y}, \qquad \frac{\partial^2 z}{\partial x^2} = -\frac{\Psi''(z)}{(\Psi'(z) - z)^3}, \tag{17.6}$$

where the primes indicate differentiation with respect to z. Thus we get

$$q = \left(\frac{y'}{x' + zy'}\right)^2 \frac{\partial z}{\partial x}, \qquad q_s = \left(\frac{y'}{x' + zy'}\right)^3 \frac{\partial^2 z}{\partial x^2} - \left(\frac{y'}{x' + zy'}\right)^4 \left(\frac{\partial z}{\partial x}\right)^2, \tag{17.7}$$

and by substituting these values in (11.1)

$$\boxed{K = \frac{\Psi''(z)}{(\Psi'(z) - y)^3} \left(\frac{y'}{x' + zy'}\right)^3}. \tag{17.8}$$

From (17.8) we see that K is then and only then a function of position only, when Ψ is linear in z. In this case we obtain from (16.20) $z = -\dfrac{x - x_0}{y - y_0}$ (x_0, y_0 constants). With regard to the theorem of §14 and to the remark on the case $z =$ const., made in §16, we have:

When, and only when, the function z which enters into the fundamental function (16.18), (16.19) of a two-dimensional Finsler space with rectilinear extremals and $I^2 = \frac{9}{2}$, has one of the values

$$z = z_0, \qquad z = -\frac{x - x_0}{y - y_0}, \qquad (x_0, y_0, z_0 \text{ constants})$$

the space is a Minkowski space. In every other case the curvature of the Finsler space depends on the line-element.

For $z = z_0$ the transformation of coordinates

$$\bar{x} = x + z_0 y, \qquad \bar{y} = y, \tag{17.9}$$

for $z = -\dfrac{x - x_0}{y - y_0}$ the transformation

$$\bar{x} = \frac{x - x_0}{y - y_0}, \qquad \bar{y} = -\frac{1}{y - y_0} \tag{17.10}$$

carries over the fundamental function in

$$F = \frac{(\bar{x}')^2}{\bar{y}'}. \tag{17.11}$$

18. Table of all two-dimensional Finsler spaces with rectilinear extremals the main scalar of which is a function of position. Now we are able to set up a table of all Finsler spaces with rectilinear extremals the main scalar I of which is a function of position only, and therefore (§15) constant. For $I^2 \neq 0, \frac{9}{2}$ these spaces are Minkowski spaces (§15). For $I = 0$, we have K constant (§15). Consequently the corresponding spaces are the Riemannian spaces of constant curvature. The case $I^2 = \frac{9}{2}$ was discussed in §§16, 17. With regard to (14.13), we obtain the following table:

(I.) $K =$ const., $\neq 0$. $I = 0$. (Non-euclidean spaces).

(1) $K = \dfrac{1}{k^2}$: $F = k\,\dfrac{[(a^2 + y^2)(x')^2 - 2xyx'y' + (a^2 + x^2)(y')^2]^{\frac{1}{2}}}{a^2 + x^2 + y^2}$,

(2) $K = -\dfrac{1}{k^2}$: $F = k\,\dfrac{[(a^2 - y^2)(x')^2 + 2xyx'y' + (a^2 - x^2)(y')^2]^{\frac{1}{2}}}{a^2 - x^2 - y^2}$;

($k > 0$; $a =$ const. > 0).

(II.) $K = 0$. $I =$ const. (Euclidean space, respectively Minkowski spaces with $I =$ const. $\neq 0$).

(3) $I^2 < 4$: $F = [(x')^2 + (y')^2]^{\frac{1}{2}} e^{\frac{I}{\sqrt{(4-I)^2}} \operatorname{arctg} \frac{y'}{x'}}$,

(4) $I^2 = 4$: $F = x' e^{\frac{y'}{x'}}$ [24]

(5) $I^2 > 4$: $F = (x')^{\frac{1}{2}\left(1 - \frac{I}{\sqrt{(I^2-4)}}\right)} (y')^{\frac{1}{2}\left(1 + \frac{I}{\sqrt{(I^2-4)}}\right)}$.

(*III*). *K variable.* $I^2 = \frac{9}{2}$.

(6) $F = \dfrac{(x' + zy')^2}{y'}$, $x + yz = \Psi(z)$, (Ψ arbitrary function, $\Psi''(z) \neq 0$).

This table shows that the two-dimensional Finsler geometries with rectilinear extremals and constant curvature $K \neq 0$ which are not Riemannian (Hilbert's geometry and its generalization, Funk's geometry of the specific metric and so on),[25] have a main scalar which depends on the line-element $(x, y; x', y')$. The same is true for the two-dimensional Finsler geometries with rectilinear extremals and $K = 0$ which are not Minkowski geometries.[25]

Finally, let us remark that the spaces, enumerated in the above table under (*I*.) and (*II*.), admit a continuous three parameter group of transformations into themselves.[26]

19. Landsberg spaces with rectilinear extremals. A *Landsberg space* is a two-dimensional Finsler space with

(19.1) $$I_{,} = 0.$$ [27]

The Landsberg spaces with rectilinear extremals are characterized by (19.1), (11.4), (11.5). In consequence of (19.1), equation (11.4) becomes

(19.2) $$I_{b\vartheta} = -2II_b\,.$$

[24] First, we have, corresponding to $I = \pm 2$, the two fundamental functions $F = x'e^{\frac{y'}{x'}}$, $F^* = x'e^{-\frac{y'}{x'}}$. The transformation of coördinates $\bar{x} = x$, $\bar{y} = -y$ transforms F^* to F, if we drop the bars.

[25] P. Funk, [9], [10].

[26] For the spaces (II), (4), (5) cf. E. Nohel, [14], (especially groups 16, 20, 25), A. Maccone, [15], E. Cartan, [5], p. 135. In these papers, the space (II), (3) does not appear, because they take as starting point the complex domain, where the types (II), (3) and (II), (5) are not different. All spaces (II), (3)-(5) figure in S. Lie and F. Engel, [13], p. 435 ff.; but their *ds* is not there given.

[27] These spaces were first considered, from another point of view, by G. Landsberg, [12], p. 334 f. See also L. Berwald, [1], pp. 208 f., 211–213; E. Cartan, [5], p. 133 f.

From (19.1) and (19.2) we obtain, with the aid of the permutation formulas (7.6), the formulas

$$
(19.3)\quad \begin{cases} I_{\vartheta b} = -II_b, & I_{\vartheta s} = -I_b, \\ I_{\vartheta\vartheta b} = (-I_\vartheta + I^2 - 1)I_b, & I_{\vartheta\vartheta s} = 3II_b, \\ I_{\vartheta\vartheta\vartheta b} = (-I_{\vartheta\vartheta} + 3II_\vartheta - I^3 + 4I)I_b, & I_{\vartheta\vartheta\vartheta s} = (4I_\vartheta - 7I^2 + 1)I_b, \\ I_{\vartheta\vartheta\vartheta\vartheta b} = (-I_{\vartheta\vartheta\vartheta} + 4II_{\vartheta\vartheta} + 3I_\vartheta^2 - 6I^2I_\vartheta + 8I_\vartheta + I^4 - 11I^2 + 1)I_b, \end{cases}
$$

which we shall use later on.

We desire to determine all Landsberg spaces with rectilinear extremals. For that purpose we state first the following theorem:

The Landsberg spaces, the extremals of which form a quasigeodesic system of curves (§12), *are identical with the two-dimensional affinely connected Finsler spaces.* (§8).

PROOF: A Landsberg space with a quasigeodesic system of extremals is characterized by (19.1), (19.2). In consequence of (19.1) the Bianchi identity (7.8) reduces to

$$(19.4)\qquad K_\vartheta = -IK.$$

The third permutation formula (7.6) gives

$$(19.5)\qquad I_{bs} = KI_\vartheta .$$

Now apply the second formula (7.6) to $\Phi = I_b$. On account of (19.2), (19.5), (19.4), it follows that

$$(19.6)\qquad I_{bb} = I_{bs\vartheta} - I_{b\vartheta s} = K(I_{\vartheta\vartheta} + II_\vartheta).$$

Further we get from the first permutation formula (7.6)

$$I_{b\vartheta b} = I_{bb\vartheta} + I_{bs} + II_{bb} .$$

When we calculate both members by means of (19.2), (19.4), (19.6), we obtain

$$(19.7)\qquad 2I_b^2 + K(I_{\vartheta\vartheta\vartheta} + 3II_{\vartheta\vartheta} + I_\vartheta^2 + 2I^2I_\vartheta + I_\vartheta) = 0.$$

Differentiating (19.7) as to ϑ and using (19.2), (19.4), we find

$$(19.8)\quad K[I_{\vartheta\vartheta\vartheta\vartheta} + 6II_{\vartheta\vartheta\vartheta} + (5I_\vartheta + 11I^2 + 1)I_{\vartheta\vartheta} + (7I_\vartheta + 6I^2 + 3)II_\vartheta] = 0.$$

If $K = 0$, it follows from (19.7) that $I_b = 0$. The space is a Minkowski space, because of $I_s = I_b = K = 0$ (§8).

Now let $K \neq 0$ and differentiate the second factor on the left of (19.8) with respect to b. With regard to (19.3), it follows that

$$(19.9)\qquad I_b(I_{\vartheta\vartheta\vartheta} + 3II_{\vartheta\vartheta} + I_\vartheta^2 + 2I^2I_\vartheta + I_\vartheta) = 0.$$

If $I_b = 0$, the space is an affinely connected space with $I =$ const., on account of (19.1) and $K \neq 0$ (§8). If

$$I_{\vartheta\vartheta\vartheta} + 3II_{\vartheta\vartheta} + I_\vartheta^2 + 2I^2I_\vartheta + I_\vartheta = 0, \tag{19.10}$$

(19.7) gives again $I_b = 0$.

Conversely, it follows from (8.9) that for a two-dimensional affinely connected Finsler space (19.1), (19.2) are satisfied.

Now it is easy to find all Landsberg spaces with rectilinear extremals. We must only take, from the table of §18, the spaces with $K \neq 0$ and add all Minkowski spaces. Hence we have the theorem:

There exist but the following types of Landsberg spaces with rectilinear extremals:

(I.) $K = 0$: Minkowski spaces.

$$F = F(x', y').$$

(II.) $K = const., \neq 0$: Non-euclidean spaces.

$$(1)\quad K = \frac{1}{k^2} \quad : \quad F = k\,\frac{[(a^2 + y^2)(x')^2 - 2xyx'y' + (a^2 + x^2)(y')^2]^{\frac{1}{2}}}{a^2 + x^2 + y^2};$$

$$(2)\quad K = -\frac{1}{k^2} \quad : \quad F = k\,\frac{[(a^2 - y^2)(x')^2 + 2xyx'y' + (a^2 - x^2)(y')^2]^{\frac{1}{2}}}{a^2 - x^2 - y^2};$$

$$(k > 0,\ a = \text{const.}, > 0).$$

(III.) K variable.

$$F = \frac{(x' + zy')^2}{y'}, \quad x + yz = \psi(z), \quad (\psi \text{ arbitrary function, } \psi''(z) \neq 0).$$

PRAGUE, BOHEMIA.

BIBLIOGRAPHY

(1) L. Berwald, *Ueber zweidimensionale allgemeine metrische Räume.* Journal f. d. reine u. ang. Math. 156 (1927), 191–222.

(2) ———, *Ueber Finslersche und Cartansche Geometrie. I. Geometrische Erklärungen der Krümmung und des Hauptskalars im zweidimensionalen Finslerschen Raum.* Mathematica, Cluj 16 (1940).

(3) ———, *Ueber Finslersche und Cartansche Geometrie. II. Invarianten bei der Variation vielfcaher Integrale und Parallelhyperflächen in Cartanschen Räumen.* Compositio Mathematica (1939).

(4) W. Blaschke und G. Bol, *Geometrie der Gewebe.* Berlin, J. Springer 1938. VIII + 339 pp.

(5) E. Cartan, *Sur un problème d'équivalence et la théorie des espaces métriques généralisés.* Mathematica, Cluj 4 (1930), 114–136.

(6) ———, *Les espaces de Finsler. Actualités scientifiques et industrielles.* 79 Paris, Hermann et Cie. 1934, 42 pp.

(7) J. Douglas, *The general geometry of paths.* Annals of Math. (2) 29 (1928), 143–168.

(8) P. Finsler, *Ueber Kurven und Flächen in allgemeinen Räumen.* Dissertation, Göttingen 1918, 120 pp.

(9) P. Funk, *Ueber Geometrien, bei denen die Geraden die Kürzesten sind.* Math. Ann. 101 (1929), 226–237.

(10) ———, *Ueber zweidimensionale Finslersche Räume, insbesondere über solche mit geradlinigen Extremalen und positiver konstanter Krümmung.* Math. Zeitschr. 40 (1935), 86-93.

(11) ———, *Contribution to the two-dimensional "Finsler's geometry"* unpublished.

(12) G. Landsberg, *Ueber die Krümmung in der Variationsrechnung.* Math. Ann. 65 (1908), 313-349.

(13) S. Lie und F. Engel, *Theorie der Transformationsgruppen III.* B. G. Teubner, Leipzig 1893. XXVII + 831 pp.

(14) E. Nohel, *Zur natürlichen Geometrie ebener Transformationsgruppen.* Sitzungsber. Akad. Wien 123 (1914), Abt. IIa., 2085-2115.

(15) A. Maccone, *Le geometrie di Sophur Lie e le geometrie non pitagoriche in generale.* Lincei Rend. Roma (5) 32^{I} (1923), 327-331.

(16) A. L. Underhill, *Invariants of the function $f(x, y, x', y')$ in the calculus of variations.* Trans. Amer. Math. Soc. 9 (1908), 316-338 (also Dissertation, Chicago 1907).

(17) H. Weyl, *Zur Infinitesimalgeometrie: Einordnung der projektiven und der konformen Auffassung.* Göttinger Nachr. 1921, 99-112.

ANNALS OF MATHEMATICS
Vol. 42, No. 1. January, 1941

ON UNITARY REPRESENTATIONS OF THE GROUP OF DE SITTER SPACE

BY L. H. THOMAS

(Received March 22, 1940)

The ten-parameter group of motions of De Sitter space is that of the real homogeneous linear transformations of w, x, y, z, and t, that leave $w^2 + x^2 + y^2 + z^2 - t^2$ invariant and that can be built up from infinitesimal transformations. It contains the sub-group of the real homogeneous linear transformations of w, x, y, and z, that leave $w^2 + x^2 + y^2 + z^2$ invariant and that can be built up from infinitesimal transformations, the six-parameter group of rotations of four-dimensional Euclidean space.

We seek differentiable representations by unitary matrices, in general infinite, of groups locally isomorphic with the ten-parameter group. Hermitian matrices are found for the infinitesimal transformations of representations which admit a transformation, perhaps to generalised matrices with a continuum of rows and columns, reducing the six-parameter sub-group to a product of its irreducible unitary representations.

1. INTRODUCTION

Wigner[1] has recently discussed the unitary representations of the inhomogeneous Lorentz group. If this group is regarded as the group of motions of a flat three plus one dimensional Riemannian space, it may be regarded as the limit for zero curvature of the group of motions of De Sitter space. Dirac[2] has considered the form that the electron wave equations take in De Sitter space. It is of some interest to discuss the unitary representations of the group of De Sitter space, which should be relevant to the quantum theory of the external properties of systems moving in such a space. These representations are investigated here by the standard methods of matrix-mechanics,[3] applied to the infinitesimal operators of the group, and it is believed that no differentiable representations admitting a transformation reducing the representation induced in the group of motions of a three-dimensional sub-space of constant positive curvature have been missed. The argument, however, lacks logical rigor in its application to the representations depending on continuous parameters.

[1] *On Unitary Representations of the Inhomogeneous Lorentz Group.* Annals of Mathematics 40 (1939) p. 145.

[2] *The Electron Wave Equation in De Sitter Space.* Annals of Mathematics 36 (1935) p. 657.

[3] See e.g. Born and Jordan. *Elementare Quantenmechanik.* Ch. III. Springer (1930).

2. The Infinitesimal Operators of the Group and their Commutation Relations

$$L = -\iota\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right),\quad M = -\iota\left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\right),\quad N = -\iota\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right);$$

$$X = -\iota\left(w\frac{\partial}{\partial x} - x\frac{\partial}{\partial w}\right),\quad Y = -\iota\left(w\frac{\partial}{\partial y} - y\frac{\partial}{\partial w}\right),\quad Z = -\iota\left(w\frac{\partial}{\partial z} - z\frac{\partial}{\partial w}\right); \tag{2.11}$$

$$U = -\iota\left(t\frac{\partial}{\partial x} - x\frac{\partial}{\partial t}\right),\quad V = -\iota\left(t\frac{\partial}{\partial y} - y\frac{\partial}{\partial t}\right),\quad W = -\iota\left(t\frac{\partial}{\partial z} - z\frac{\partial}{\partial t}\right),$$

and

$$T = -\iota\left(t\frac{\partial}{\partial w} - w\frac{\partial}{\partial t}\right); \tag{2.12}$$

are a complete set of ten self-adjoint linear differential operators of the group, satisfying the 45 commutation relations:—

$$\begin{aligned}
&MN - NM = \iota L,\quad MZ - ZM = \iota X,\quad YN - NY = \iota X,\\
&\qquad LX - XL = 0,\quad YZ - ZY = \iota L,\\
&NL - LN = \iota M,\quad NX - XN = \iota Y,\quad ZL - LZ = \iota Y,\\
&\qquad MY - YM = 0,\quad ZX - XZ = \iota M,\\
&LM - ML = \iota N,\quad LY - YL = \iota Z,\quad XM - MX = \iota Z,\\
&\qquad NZ - ZN = 0,\quad XY - YX = \iota N;
\end{aligned} \tag{2.21}$$

$$\begin{aligned}
&MW - WM = \iota U\quad VN - NV = \iota U,\quad LU - UL = 0,\\
&\qquad YW - WY = 0,\quad VZ - ZV = 0,\\
&NU - UN = \iota V,\quad WL - LW = \iota V,\quad MV - VM = 0,\\
&\qquad ZU - UZ = 0,\quad WX - XW = 0,\\
&LV - VL = \iota W,\quad UM - MU = \iota W,\quad NW - WN = 0,\\
&\qquad XV - VX = 0,\quad UY - YU = 0,
\end{aligned}$$

$$\begin{aligned}
&UX - XU = \iota T,\quad LT - TL = 0,\quad XT - TX = \iota U,\\
&VY - YV = \iota T,\quad MT - TM = 0,\quad YT - TY = \iota V,\\
&WZ - ZW = \iota T,\quad NT - TN = 0,\quad ZT - TZ = \iota W;
\end{aligned} \tag{2.22}$$

$$\begin{aligned} UT - TU &= \iota X, & VW - WV &= -\iota L, \\ VT - TV &= \iota Y, & WU - UW &= -\iota M, \\ WT - TW &= \iota Z, & UV - VU &= -\iota N. \end{aligned} \qquad 2.23$$

The first six of these operators (2.11) satisfying the first fifteen commutation relations (2.21) are a complete set of operators of the six-parameter sub-group.

To any differentiable representation of the ten-parameter group there must correspond a representation of these operators. If the representation of the group is unitary, that of these operators is Hermitian. A transformation reducing a representation of the sub-group to a sum of its irreducible unitary representations must reduce the representation of its operators to a sum of irreducible Hermitian representations.

The commutation relations of the displacement operators for free motion in quantum theory in De Sitter space can be reduced to the above form by change of phase.

3. Outline of the Method Employed

Let the (known) representations of the operators, $A = L, M, N, X, Y$, and Z, of the sub-group be expressed in the form (in Dirac's notation)

$$(\alpha \mid A(\beta) \mid \alpha') \qquad 3.1$$

where β is a set of parameters taking a denumerable infinity of discrete values, one for each irreducible representation, and α, α', are sets of parameters numbering the columns and rows of the matrices, and taking for each β only a finite number of discrete values. Then a transformation reducing a representation of the sub-group to a sum of these representations reduces it to the form

$$\delta_{\gamma\gamma'}\delta_{\beta\beta'} \quad (\alpha \mid A(\beta) \mid \alpha') \qquad 3.2$$

where $\delta_{\beta\beta'}$ is the usual Kronecker δ; γ, γ', are sets of parameters numbering the columns and rows of the (generalised) matrices, taking for each β, perhaps a continuous, perhaps a discrete set of values, perhaps no values at all, and $\delta_{\gamma\gamma'}$ is to be interpreted accordingly.

If the remaining operators, $P = U, V, W$, and T, of the group admit the transformation, they take the form

$$(\gamma, \beta, \alpha \mid P \mid \gamma', \beta', \alpha'). \qquad 3.3$$

The 24 commutation relations of the P's with the A's (2.22) enable us to express the P's in the form

$$(\gamma \mid P(\beta, \beta') \mid \gamma')(\alpha \mid P(\beta, \beta') \mid \alpha') \qquad 3.4$$

where $(\alpha \mid P(\beta, \beta') \mid \alpha')$ are determined, and indeed for any β fail to vanish only for at most four values of β'.

In virtue of the fifteen commutation relations of the A's (2.21) and the 24 commutation relations of the P's with the A's (2.22), the remaining six commuta-

tion relations of the P's (2.23) follow from any one of them. This one then determines possible forms for $(\gamma \mid P(\beta, \beta') \mid \gamma')$.

4. The Representations of the Sub-group

The work is shortened if the sub-group is expressed as the direct product of two three-parameter groups.

Let
$$\begin{aligned} A_1 &= L + X, & A_2 &= L - X, \\ B_1 &= M + Y, & B_2 &= M - Y, \\ C_1 &= N + Z, & C_2 &= N - Z, \end{aligned} \tag{4.1}$$

so that any one of A_1, B_1, and C_1, commutes with any one of A_2, B_2, and C_2, and these are the operators of factor groups for either of which

$$\begin{aligned} BC - CB &= 2\iota A, \\ CA - AC &= 2\iota B, \\ AB - BA &= 2\iota C. \end{aligned} \tag{4.2}$$

If
$$\begin{aligned} A - \iota B &= C_-, \\ A + \iota B &= C_+, \end{aligned}$$

we have
$$\begin{aligned} CC_- - C_-C &= -2C_-, \\ CC_+ - C_+C &= 2C_+, \\ C_-C_+ - C_+C_- &= -4C, \end{aligned} \tag{4.3}$$

and the only non-zero components of an irreducible (Hermitian) representation may be brought to the form

$$\begin{aligned} (m \mid C_- \mid m + 1) &= 2\sqrt{\{(j - m)(j + m + 1)\}}, \\ & \qquad m = -j, -j + 1, \cdots, j - 1, j, \\ (m \mid C_+ \mid m - 1) &= 2\sqrt{\{(j + m)(j - m + 1)\}}, \\ (m \mid C \mid m) &= 2m, \end{aligned} \tag{4.4}$$

where for different representations

$$j = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots . \tag{4.5}$$

Thus we may take for β above j_1 and j_2 for these factor groups, taking any values from (4.5), and for α above m_1 and m_2 satisfying (4.4), and the matrices representing L, M, N, X, Y, and Z, may be written down immediately, in particular

$$(m_1, m_2 \mid Z(j_1, j_2) \mid m_1, m_2) = m_1 - m_2 \tag{4.6}$$

for $m_1 = -j_1, -j_1 + 1, \cdots, j_1$, $m_2 = -j_2, -j_2 + 1, \cdots, j_2$, the remaining components vanishing.

5. The Determination of $(\alpha \mid P(\beta, \beta') \mid \alpha')$

The 24 commutation relations (2.22) give for the operators of the first factor group

$$(T + \iota W)C - C(T + \iota W) = -(T + \iota W), \qquad (U - \iota V)C - C(U - \iota V) = (U - \iota V), \tag{5.11}$$

$$(T + \iota W)C_+ - C_+(T + \iota W) = 0, \qquad (U - \iota V)C_- - C_-(U - \iota V) = 0, \tag{5.12}$$

$$(T + \iota W)C_- - C_-(T + \iota W) = -2\iota(U - \iota V), \qquad (U - \iota V)C_+ - C_+(U - \iota V) = 2\iota(U - \iota V), \tag{5.13}$$

and similar equations for $T - \iota W$ and $U - \iota V$ with C_+ and C_- interchanged and the sign of C reversed; and for the second factor group similar equations with the sign of T reversed.

For (5.11) to hold it is necessary and sufficient that a component of the transformed representation of $T + \iota W$ vanish unless $m' = m - \frac{1}{2}$, one of $U - \iota V$ unless $m' = m + \frac{1}{2}$, and then for (5.12) to hold that in these cases they involve m only in factors $\sqrt{\{(j + m)!(j' - m + \frac{1}{2})!/(j - m)!/(j' + m - \frac{1}{2})!\}}$ and $\sqrt{\{(j - m)!(j' + m + \frac{1}{2})!/(j + m)!(j' - m - \frac{1}{2})!\}}$ respectively. (5.13) gives then

$$\{(j - j')^2 - \tfrac{1}{4}\}\{(j + j' + 1)^2 - \tfrac{1}{4}\} = 0,$$

so that the components vanish unless $j' = j - \frac{1}{2}$ or $j' = j + \frac{1}{2}$, and we must have for the only non-zero components,

$$(\lambda, j_1, m_1 \mid T + \iota W \mid \lambda', j_1 - \tfrac{1}{2}, m_1 - \tfrac{1}{2}) = (\lambda \mid u(j_1) \mid \lambda')\sqrt{(j_1 + m_1)},$$

$$(\lambda, j_1, m_1 \mid T + \iota W \mid \lambda', j_1 + \tfrac{1}{2}, m_1 - \tfrac{1}{2}) = (\lambda \mid v(j_1) \mid \lambda')\sqrt{(j_1 - m_1 + 1)},$$

$$(\lambda, j_1, m_1 \mid U - \iota V \mid \lambda', j_1 - \tfrac{1}{2}, m_1 + \tfrac{1}{2}) = -\iota(\lambda \mid u(j_1) \mid \lambda')\sqrt{(j_1 - m_1)},$$

$$(\lambda, j_1, m_1 \mid U - \iota V \mid \lambda', j_1 + \tfrac{1}{2}, m_1 + \tfrac{1}{2}) = \iota(\lambda \mid v(j_1) \mid \lambda')\sqrt{(j_1 + m_1 + 1)},$$

where λ includes γ, j_2, and m_2; necessary and sufficient conditions that (5.11), (5.12), and (5.13), be satisfied, $u(j_1)$ and $v(j_1)$ being arbitrary matrix functions of j_1 for $j_1 = \frac{1}{2}, 1, \frac{3}{2}, \cdots$, and for $j_1 = 0, \frac{1}{2}, 1, \cdots$, respectively.

Similarly, for their only non-zero components,

$$(\lambda, j_1, m_1 \mid T - \iota W \mid \lambda', j_1 + \tfrac{1}{2}, m_1 + \tfrac{1}{2}) = (\lambda \mid u'(j_1) \mid \lambda')\sqrt{(j_1 + m_1 + 1)},$$

$$(\lambda, j_1, m_1 \mid T - \iota W \mid \lambda', j_1 - \tfrac{1}{2}, m_1 + \tfrac{1}{2}) = (\lambda \mid v'(j_1) \mid \lambda')\sqrt{(j_1 - m_1)},$$

$$(\lambda, j_1, m_1 \mid U + \iota V \mid \lambda', j_1 + \tfrac{1}{2}, m_1 - \tfrac{1}{2}) = \iota(\lambda \mid u'(j_1) \mid \lambda')\sqrt{(j_1 - m_1 + 1)},$$

$$(\lambda, j_1, m_1 \mid U + \iota V \mid \lambda', j_1 - \tfrac{1}{2}, m_1 - \tfrac{1}{2}) = -\iota(\lambda \mid v'(j_1) \mid \lambda')\sqrt{(j_1 + m_1)},$$

where $u'(j_1)$ and $v'(j_1)$ are further arbitrary matrix functions of j_1 for $j_1 = 0, \frac{1}{2}, 1, \cdots$, and for $j_1 = \frac{1}{2}, 1, \frac{3}{2}, \cdots$, respectively.

Combining these results with the corresponding results for the second factor group, for which the components of $T + \iota W$ are connected with those of $U + \iota V$, and those of $T - \iota W$ with those of $U - \iota V$, we obtain the necessary and sufficient conditions that the 24 commutation relations (2.22) be satisfied with L, M, N, X, Y, and Z, given as in (4.). The only non-zero components of $T + \iota W$, $U - \iota V$, $U + \iota V$, and $T - \iota W$, have the form

$$
\begin{aligned}
&(\gamma, j_1, m_1, j_2, m_2 \mid T + \iota W \mid \gamma', j_1 - \tfrac{1}{2}, m_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}, m_2 + \tfrac{1}{2}) \\
&\qquad = (\gamma \mid f(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 + m_1)(j_2 - m_2)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid T + \iota W \mid \gamma', j_1 + \tfrac{1}{2}, m_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}, m_2 + \tfrac{1}{2}) \\
&\qquad = (\gamma \mid g(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 - m_1 + 1)(j_2 - m_2)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid T + \iota W \mid \gamma', j_1 - \tfrac{1}{2}, m_1 - \tfrac{1}{2}, j_2 + \tfrac{1}{2}, m_2 + \tfrac{1}{2}) \\
&\qquad = (\gamma \mid h(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 + m_1)(j_2 + m_2 + 1)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid T + \iota W \mid \gamma', j_1 + \tfrac{1}{2}, m_1 - \tfrac{1}{2}, j_2 + \tfrac{1}{2}, m_2 + \tfrac{1}{2}) \\
&\qquad = (\gamma \mid k(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 - m_1 + 1)(j_2 + m_2 + 1)\}}.
\end{aligned}
\tag{5.2}
$$

$$
\begin{aligned}
&(\gamma, j_1, m_1, j_2, m_2 \mid U - \iota V \mid \gamma', j_1 - \tfrac{1}{2}, m_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2}, m_2 + \tfrac{1}{2}) \\
&\qquad = -\iota(\gamma \mid f(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 - m_1)(j_2 - m_2)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid U - \iota V \mid \gamma', j_1 + \tfrac{1}{2}, m_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2}, m_2 + \tfrac{1}{2}) \\
&\qquad = \iota(\gamma \mid g(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 + m_1 + 1)(j_2 - m_2)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid U - \iota V \mid \gamma', j_1 - \tfrac{1}{2}, m_1 + \tfrac{1}{2}, j_2 + \tfrac{1}{2}, m_2 + \tfrac{1}{2}) \\
&\qquad = -\iota(\gamma \mid h(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 - m_1)(j_2 + m_2 + 1)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid U - \iota V \mid \gamma', j_1 + \tfrac{1}{2}, m_1 + \tfrac{1}{2}, j_2 + \tfrac{1}{2}, m_2 + \tfrac{1}{2}) \\
&\qquad = \iota(\gamma \mid k(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 + m_1 + 1)(j_2 + m_2 + 1)\}},
\end{aligned}
$$

$$
\begin{aligned}
&(\gamma, j_1, m_1, j_2, m_2 \mid U + \iota V \mid \gamma', j_1 - \tfrac{1}{2}, m_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}, m_2 - \tfrac{1}{2}) \\
&\qquad = \iota(\gamma \mid f(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 + m_1)(j_2 + m_2)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid U + \iota V \mid \gamma', j_1 + \tfrac{1}{2}, m_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}, m_2 - \tfrac{1}{2}) \\
&\qquad = \iota(\gamma \mid g(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 - m_1 + 1)(j_2 + m_2)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid U + \iota V \mid \gamma', j_1 - \tfrac{1}{2}, m_1 - \tfrac{1}{2}, j_2 + \tfrac{1}{2}, m_2 - \tfrac{1}{2}) \\
&\qquad = -\iota(\gamma \mid h(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 + m_1)(j_2 - m_2 + 1)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid U + \iota V \mid \gamma', j_1 + \tfrac{1}{2}, m_1 - \tfrac{1}{2}, j_2 + \tfrac{1}{2}, m_2 - \tfrac{1}{2}) \\
&\qquad = -\iota(\gamma \mid k(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 - m_1 + 1)(j_2 - m_2 + 1)\}},
\end{aligned}
$$

$$
\begin{aligned}
&(\gamma, j_1, m_1, j_2, m_2 \mid T - \iota W \mid \gamma', j_1 - \tfrac{1}{2}, m_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2}, m_2 - \tfrac{1}{2}) \\
&\qquad = -(\gamma \mid f(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 - m_1)(j_2 + m_2)\}}, \\
&(\gamma, j_1, m_1, j_2, m_2 \mid T - \iota W \mid \gamma', j_1 + \tfrac{1}{2}, m_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2}, m_2 - \tfrac{1}{2}) \\
&\qquad = (\gamma \mid g(j_1, j_2) \mid \gamma')\sqrt{\{(j_1 + m_1 + 1)(j_2 + m_2)\}},
\end{aligned}
$$

$$(\gamma, j_1, m_1, j_2, m_2 \mid T - \iota W \mid \gamma', j_1 - \tfrac{1}{2}, m_1 + \tfrac{1}{2}, j_2 + \tfrac{1}{2}, m_2 - \tfrac{1}{2}) = (\gamma \mid h(j_1, j_2) \mid \gamma' \sqrt{\{(j_1 - m_1)(j_2 - m_2 + 1)\}},$$

$$(\gamma, j_1, m_1, j_2, m_2 \mid T - \iota W \mid \gamma', j_1 + \tfrac{1}{2}, m_1 + \tfrac{1}{2}, j_2 + \tfrac{1}{2}, m_2 - \tfrac{1}{2}) = -(\gamma \mid k(j_1, j_2) \mid \gamma') \sqrt{\{(j_1 + m_1 + 1)(j_2 - m_2 + 1)\}},$$

where $f(j_1, j_2)$, $g(j_1, j_2)$, $h(j_1, j_2)$, and $k(j_1, j_2)$, are in general arbitrary matrix functions of j_1 and j_2, j_1 taking values from $\frac{1}{2}$, 1, $\frac{3}{2}$, $\cdots$, in f and h, from 0, $\frac{1}{2}$, 1, $\cdots$, in g and k, j_2 taking values from $\frac{1}{2}$, 1, $\frac{3}{2}$, $\cdots$, in f and g, from 0, $\frac{1}{2}$, 1, $\cdots$, in h and k, but not necessarily all such pairs of values.

If the representation is Hermitian, $f(j_1, j_2)$ and $-k(j_1 - \frac{1}{2}, j_2 - \frac{1}{2})$ must be Hermitian conjugates, and $g(j_1, j_2)$ and $h(j_1 + \frac{1}{2}, j_2 - \frac{1}{2})$ must be Hermitian conjugates. 5.3

It should be noted that the factors involving m_1 and m_2 make components connecting allowed values of m_1 and m_2 to those that are not allowed vanish, so that we need not usually take explicit consideration of the ranges of m_1 and m_2.

6. The General Form of f, g, h, and k

The remaining six commutation relations now reduce to any one of them; in particular $WT - TW = \iota Z$ in the form

$$(T + \iota W)(T - \iota W) - (T - \iota W)(T + \iota W) = -2Z$$

gives

$$\begin{aligned}
j_1 f(j_1, j_2) g(j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}) &= (j_1 + 1) g(j_1, j_2) f(j_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2}),\\
j_2 f(j_1, j_2) h(j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}) &= (j_2 + 1) h(j_1, j_2) f(j_1 - \tfrac{1}{2}, j_2 + \tfrac{1}{2}),\\
j_2 g(j_1, j_2) k(j_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2}) &= (j_2 + 1) k(j_1, j_2) g(j_1 + \tfrac{1}{2}, j_2 + \tfrac{1}{2}),\\
j_1 h(j_1, j_2) k(j_1 - \tfrac{1}{2}, j_2 + \tfrac{1}{2}) &= (j_1 + 1) k(j_1, j_2) h(j_1 + \tfrac{1}{2}, j_2 + \tfrac{1}{2}),
\end{aligned} \qquad 6.11$$

for $j_1' = j_1, j_2' = j_2 - 1$; $j_1' = j_1 - 1, j_2' = j_2$; $j_1' = j_1 + 1, j_2' = j_2$; and $j_1' = j_1$, $j_2' = j_2 + 1$; respectively in β', β: and

$$\begin{aligned}
&(m_1 j_2 - m_2 j_1) f(j_1, j_2) k(j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2})\\
&\quad + (m_1 j_2 + m_2(j_1 + 1)) g(j_1, j_2) h(j_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2})\\
&\quad - (m_1(j_2 + 1) + m_2 j_1) h(j_1, j_2) g(j_1 - \tfrac{1}{2}, j_2 + \tfrac{1}{2})\\
&\quad - (m_1(j_2 + 1) - m_2(j_1 + 1)) k(j_1, j_2) f(j_1 + \tfrac{1}{2}, j_2 + \tfrac{1}{2}) = m_1 - m_2,
\end{aligned} \qquad 6.12$$

for $j_1' = j_1$, $j_2' = j_2$: all the other components vanishing identically.

These relations hold whenever both j_1, j_2, and j_1', j_2', are possible values of β, non-existent terms being interpreted as vanishing.

Irreducible solutions of these equations are of two types; those for which the least value of either j_1 or j_2 in β in any non-vanishing component of (3.3) differs from zero; and those in which both these values are zero.

For the first type suppose, for example, $j_1 \nless p > 0$, and $j_1 = p$, for $j_2 = q$. Then (6.12) gives two independent relations between non-vanishing components of $g(p, q)h(p + \frac{1}{2}, q - \frac{1}{2})$ and $k(p, q)f(p + \frac{1}{2}, q + \frac{1}{2})$, which are

therefore constant multiples of an idempotent matrix, and $g(p, q)$ commutes with $h(p + \frac{1}{2}, q - \frac{1}{2})$, and $k(p, q)$ with $f(p + \frac{1}{2}, q + \frac{1}{2})$. Equations (6.11) then show that $g(p, q + 1)h(p + \frac{1}{2}, q + \frac{1}{2})$, $k(p, q - 1)f(p, q - \frac{1}{2})$, $g(p + \frac{1}{2}, q + \frac{1}{2})h(p + 1, q)$, and $k(p + \frac{1}{2}, q - \frac{1}{2})f(p + 1, q)$, are sums of multiples, zero for the first two, of the same idempotent matrix and of matrices whose product with it vanishes. Equations (6.12) then give similar information about $g(p + \frac{1}{2}, q - \frac{1}{2})h(p + 1, q - 1)$, $g(p + 1, q)h(p + \frac{3}{2}, q - \frac{1}{2})$, $k(p + 1, q)f(p + \frac{3}{2}, q + \frac{1}{2})$, and $k(p + \frac{1}{2}, q + \frac{1}{2})f(p + 1, q + 1)$; and so on. There are more than sufficient equations. If inconsistent they show that no such representation exists; if consistent that it decomposes into a term for which, for non-vanishing components, $j_1 > p$ for $j_2 = q$ which can be treated in the same way, and a term for which each $g(j_1, j_2)h(j_1 + \frac{1}{2}, j_2 - \frac{1}{2})$ and $k(j_1, j_2)f(j_1 + \frac{1}{2}, j_2 + \frac{1}{2})$ is a determined multiple of the same idempotent matrix.

This result and equations (6.11) now show that different transformations among γ, γ', for various j_1, j_2, can be taken so as to make all of f, g, h, and k, simultaneously diagonal, and the representation is reduced to a sum of representations in which f, g, h, and k, have but one row and column; if the original representation was Hermitian, the transformations required form a unitary transformation.

For the second type suppose $j_1 = 0$ for $j_2 = q$.

(6.12) now gives only one relation between $g(0, q)h(\frac{1}{2}, q - \frac{1}{2})$ and $k(0, q)f(\frac{1}{2}, q + \frac{1}{2})$; if we make one of these diagonal, so must be the other, and the argument for the previous case, introducing at each stage different transformations among γ, γ', for various j_1, j_2, shows that either no representation exists, or that it decomposes into a term for which always $j_1 > 0$, for $j_2 = q$, and a sum of terms for each of which each $g(j_1, j_2)h(j_1 + \frac{1}{2}, j_2 - \frac{1}{2})$ and $k(j_1, j_2)f(j_1 + \frac{1}{2}, j_2 + \frac{1}{2})$ is a determined multiple of an idempotent matrix, the products of the several idempotent matrices vanishing; perhaps continuously many such terms corresponding to possibly continuously many characteristic values of $g(0, q)h(\frac{1}{2}, q - \frac{1}{2})$. Finally as in the previous case, the representation reduces to a sum of representations in which f, g, h, and k, have but one row and column. It is this argument, involving matrices with continuously many rows and columns, which lacks logical rigor.

Thus for irreducible representations, $f(j_1, j_2)$, $g(j_1, j_2)$, $h(j_1, j_2)$, and $k(j_1, j_2)$, in (5.1) can be treated as numbers, $f(j_1, j_2)$, $g(j_1, j_2)$, $h(j_1, j_2)$, and $k(j_1, j_2)$. 6.2

Moreover the relations (6.11) and (6.12) involve either only values of β for which $j_1 + j_2$ is an integer or only values for which it is half an odd integer, so that any irreducible representation involves either only integral values of $j_1 + j_2$ or only half odd integral values. 6.3

7. The Solution of the Recurrence Relations

Equations (6.11), if $g(j_1, j_2)g(j_1 - \frac{1}{2}, j_2 - \frac{1}{2})$ and $h(j_1, j_2 - 1)h(j_1 + \frac{1}{2}, j_2 - \frac{1}{2})$ do not both vanish, give

$$j_1(j_2 + \tfrac{1}{2})f(j_1, j_2)k(j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2})$$
$$= (j_1 + 1)(j_2 - \tfrac{1}{2})f(j_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2})k(j_1, j_2 - 1),$$

so that in general

$$f(j_1, j_2)k(j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}) = -\frac{F(j_1 + j_2)}{j_1(j_1 + \frac{1}{2})j_2(j_2 + \frac{1}{2})}. \qquad 7.11$$

Likewise, unless $f(j_1, j_2)f(j_1 + \frac{1}{2}, j_2 - \frac{1}{2})$ and $k(j_1 - \frac{1}{2}, j_2 - \frac{1}{2})k(j_1, j_2 - 1)$ both vanish for values of j_1 and j_2 involved,

$$g(j_1, j_2)h(j_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2}) = \frac{G(j_1 - j_2)}{(j_1 + \frac{1}{2})(j_1 + 1)j_2(j_2 + \frac{1}{2})} \qquad 7.12$$

Then (6.12) gives

$$\frac{(m_1j_2 - m_2j_1)F(j_1 + j_2)}{j_1(j_1 + \frac{1}{2})j_2(j_2 + \frac{1}{2})} - \frac{(m_1(j_2 + 1) - m_2(j_1 + 1))F(j_1 + j_2 + 1)}{(j_1 + \frac{1}{2})(j_1 + 1)(j_2 + \frac{1}{2})(j_2 + 1)}$$
$$- \frac{(m_1j_2 + m_2(j_1 + 1))G(j_1 - j_2)}{(j_1 + \frac{1}{2})(j_1 + 1)j_2(j_2 + \frac{1}{2})} + \frac{(m_1(j_2 + 1) - m_2(j_1 + 1))G(j_1 + j_2 + 1)}{j_1(j_1 + \frac{1}{2})(j_2 + \frac{1}{2})(j_2 + 1)} \qquad 7.2$$
$$= m_2 - m_1$$

Putting $m_2 = m_1$ in (7.2) we obtain

$$\frac{F(j_1 + j_2) + G(j_1 + j_2)}{j_2(j_1 + \frac{1}{2})} + \frac{F(j_1 + j_2 + 1) + G(j_1 - j_2 - 1)}{(j_2 + 1)(j_1 + \frac{1}{2})}$$
$$= \frac{F(j_1 + j_2) + G(j_1 - j_2 - 1)}{j_1(j_2 + \frac{1}{2})} + \frac{F(j_1 + j_2 + 1) + G(j_1 - j_2)}{(j_1 + 1)(j_2 + \frac{1}{2})},$$

from which follows

$$F(j_1 + j_2) + G(j_1 - j_2) = j_2(j_1 + \tfrac{1}{2})[\varphi(j_1 + j_2) + \psi(j_1 - j_2)]. \qquad 7.3$$

Subtracting this equation from the corresponding equation with j_1 and j_2 replaced by $j_1 + \frac{1}{2}$ and $j_2 + \frac{1}{2}$ we see that ψ is a quadratic function. Likewise φ is a quadratic function, so that F and G are of the fourth degree, and for (7.3) to hold, of the forms

$$F(j_1 + j_2) = a[(j_1 + j_2)^2 + (j_1 + j_2)]^2 + b[(j_1 + j_2)^2 + (j_1 + j_2)] + c,$$
$$G(j_1 - j_2) = -a[(j_1 - j_2)^2 + (j_1 - j_2)]^2 - b[(j_1 - j_2)^2 + (j_1 - j_2)] - c.$$

Substituting these values, (7.2) is satisfied identically if $a = 1/16$, and we have

$$-f(j_1, j_2)k(j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2})$$
$$= \frac{(j_1 + j_2 - p)(j_1 + j_2 + p + 1)(j_1 + j_2 - q)(j_1 + j_2 + q + 1)}{2j_1(2j_1 + 1)2j_2(2j_2 + 1)}, \qquad 7.41$$

$$g(j_1, j_2)k(j_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2})$$
$$= -\frac{(j_1 - j_2 - p)(j_1 - j_2 + p + 1)(j_1 - j_2 - q)(j_1 - j_2 + q + 1)}{(2j_1 + 1)(2j_1 + 2)2j_2(2j_2 + 1)}, \qquad 7.42$$

where p and q are arbitrary.

Starting with any given pair of values of j_1 and j_2 from $0, \frac{1}{2}, \frac{3}{2}, \cdots$, for which these products do not vanish, we may use the recurrence relations to find the values of these products for adjoining pairs of such values for which the sum of j_1 and j_2 has the initial form (either an integer or half an odd integer) unless stopped by zero factors in the numerators, which must occur before reaching pairs of values outside the above range. In this case the separate factors of these products can always be chosen so that all the equations (6.11) are satisfied, and (5.1) yield matrices satisfying all the commutation relations. 7.5

In the special cases in which either all the products $f(j_1, j_2)k(j_1 - \frac{1}{2}, j_2 - \frac{1}{2})$ or all the products $g(j_1, j_2)h(j_1 + \frac{1}{2}, j_2 - \frac{1}{2})$ vanish, it can be verified that the surviving recurrence relations suffice to require the same form (7.41) or (7.42) for the surviving products.

8. Invariants

For matrices of the above form

$$T^2 + U^2 + V^2 + W^2 - (L^2 + M^2 + N^2 + X^2 + Y^2 + Z^2) \qquad 8.1$$

must exist, and, commuting with each matrix, be an invariant.

Writing it in the form

$$\begin{aligned}\tfrac{1}{2}[(T + \iota W)(T - \iota W) + (T - \iota W)(T + \iota W)]& \\ + \tfrac{1}{2}[(U + \iota V)(U - \iota V) + (U - \iota V)(U + \iota V)]& \\ - \tfrac{1}{2}[(L + X)^2 + (M + Y)^2 + (N + Z)^2]& \\ - \tfrac{1}{2}[(L - X)^2 + (M - Y)^2 + (N - Z)^2]&,\end{aligned}$$

we see that unless $\alpha' = \alpha$ and $\beta' = \beta$, its components vanish, and that for β and β' given by j_1, j_2, the (diagonal) component has the form

$$\begin{aligned}-2j_1j_2f(j_1, j_2)k(j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}) + 2(j_1 + 1)j_2g(j_1, j_2)h(j_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2})& \\ + 2j_1(j_2 + 1)h(j_1, j_2)g(j_1 - \tfrac{1}{2}, j_2 + \tfrac{1}{2})& \\ - 2(j_1 + 1)(j_2 + 1)k(j_1, j_2)f(j_1 + \tfrac{1}{2}, j_2 + \tfrac{1}{2})& \qquad 8.11 \\ - 2j_1(j_1 + 1) - 2j_2(j_2 + 1)& \\ = \tfrac{5}{2} - (q + \tfrac{1}{2})^2 - (p + \tfrac{1}{2})^2&,\end{aligned}$$

which must be its value for a representation determined by q and p.

Again

$$\begin{aligned}(LX + MY + NZ)^2 - (LU + MV + NW)^2 - (LT + WY - UZ)^2& \\ - (MT + UZ - WX)^2 - (NT + VX - UY)^2&\end{aligned} \qquad 8.2$$

must exist, and, commuting with each matrix, be an invariant.

Writing this in the form

$$K^2 + (KT - TK)^2 + (KU - UK)^2 + (KV - VK)^2 + (KW - WK)^2,$$

where $K = LX + MY + NZ$, has only diagonal components $j_1(j_1 + 1) - j_2(j_2 + 1)$, so that we can at once write down the components of $K(T + \iota W) - (T + \iota W)K$, etc., we see that unless $\alpha = \alpha'$ and $\beta = \beta'$, the components of (8.2) vanish, and that for β and β' given by j_1, j_2, the diagonal component has the form

$$\begin{aligned}(j_1(j_1 + 1) - j_2(j_2 + 1))^2 &+ 2j_1j_2(j_1 - j_2)^2 f(j_1, j_2)k(j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}) \\ &- 2(j_1 + 1)j_2(j_1 + j_2 + 1)^2 g(j_1, j_2)h(j_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2}) \\ &- 2j_1(j_2 + 1)(j_1 + j_2 + 1)^2 h(j_1, j_2)g(j_1 - \tfrac{1}{2}, j_2 + \tfrac{1}{2}) \\ &+ 2(j_1 + 1)(j_2 + 1)(j_1 - j_2)^2 k(j_1, j_2)f(j_1 + \tfrac{1}{2}, j_2 + \tfrac{1}{2}) \\ &= p(p + 1)q(q + 1),\end{aligned} \tag{8.21}$$

which must be its value for a representation determined by q and p.

These results enable us to show that representations for distinct pairs of values of q and p are distinct. (In the quantum theory of the free motion of a system in De Sitter space, (8.1) would be proportional to the square of the rest mass, (8.2) for given mass, to the negative square of the spin.)

9. The Allowed Values of p and q

In order that the matrices should be Hermitian, $f(j_1, j_2)$ and $-k(j_1 - \frac{1}{2}, j_2 - \frac{1}{2})$ must be complex conjugates and $g(j_1, j_2)$ and $k(j_1 + \frac{1}{2}, j_2 + \frac{1}{2})$ must be complex conjugates, (5.3), (6.2). Thus for all pairs of values taken by j_1, j_2, we must have

$$\begin{aligned}-f(j_1, j_2)k(j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}) &\not< 0, \\ g(j_1, j_2)h(j_1 + \tfrac{1}{2}, j_2 - \tfrac{1}{2}) &\not< 0.\end{aligned} \tag{9.1}$$

Then by altering the phases, we can make

$$\begin{aligned} f(j_1, j_2) &= \iota\sqrt{\left\{\frac{(j_1 + j_2 - p)(j_1 + j_2 + p + 1)(j_1 + j_2 - q)(j_1 + j_2 + q + 1)}{2j_1(2j_1 + 1)2j_2(2j_2 + 1)}\right\}}, \\ g(j_1, j_2) &= \sqrt{\left\{-\frac{(j_1 - j_2 - p)(j_1 - j_2 + p + 1)(j_1 - j_2 - q)(j_1 - j_2 + q + 1)}{(2j_1 + 1)(2j_2 + 2)2j_2(2j_2 + 1)}\right\}}, \\ h(j_1, j_2) &= \sqrt{\left\{-\frac{(j_1 - j_2 - p - 1)(j_1 - j_2 + p)(j_1 - j_2 - q - 1)(j_1 - j_2 + q)}{2j_1(2j_1 + 1)(2j_2 + 1)(2j_2 + 2)}\right\}}, \\ k(j_1, j_2) &= \iota\sqrt{\left\{\frac{(j_1 + j_2 - p + 1)(j_1 + j_2 + p + 2)(j_1 + j_2 - q + 1)(j_1 + j_2 + q + 2)}{(2j_1 + 1)(2j_1 + 2)(2j_2 + 1)(2j_2 + 2)}\right\}}, \end{aligned} \tag{9.2}$$

where the radicals are taken real and positive or zero.

Either I only integral values of $j_1 + j_2$, and therefore also of $j_1 - j_2$, are involved, or II only half odd integral values, (6.3).

I. For factors in the numerators to vanish (7.5), at least one of p, q, $-p - 1$, or $-q - 1$, must be a positive integer or zero; since interchange of p and q,

or of p and $-p - 1$ does not affect (9.2), we may suppose $p \nless 0$ is an integer. Then $(q + \frac{1}{2})^2$ must be real for the expressions under the radical signs to be real.

IA. Suppose q is not an integer. Since a zero factor in the numerator of $f(j_1, j_2)$ must occur at the lower limit of $j_1 + j_2$ which must be not less than zero, $j_1 + j_2$ must take in $f(j_1, j_2)$ the values $p + 1, p + 2, \cdots$, and likewise in $k(j_1, j_2)$ the values $p, p + 1, \cdots$, which must make the radicals real and positive.

Thus

$$(p + \tfrac{3}{2})^2 > (q + \tfrac{1}{2})^2. \tag{9.3}$$

$j_1 + j_2$ then takes in $g(j_1, j_2)$ and $h(j_1, j_2)$ the values $p, p + 1, \cdots$.

Since a zero factor in the numerator of $g(j_1, j_2)$ must occur, for any allowed value of $j_1 + j_2$, in particular $j_1 + j_2 = p$, at lower and upper limits of $j_1 - j_2$ which must make both j_1 and j_2 positive or zero, $j_1 - j_2$ must take in $g(j_1, j_2)$ the values $-p, -p + 1, \cdots, p - 1$, and likewise in $h(j_1, j_2)$ the values $-p + 1, -p + 2, \cdots, p$, and in $f(j_1, j_2)$ and $k(j_1, j_2)$ the values $-p$, $-p + 1, \cdots, p$.

For $p > 0$, therefore, $j_1 - j_2 = 0$ being a possible value in $g(j_1, j_2)$ or $h(j_1, j_2)$,

$$(\tfrac{1}{2})^2 > (q + \tfrac{1}{2})^2, \tag{9.41}$$

which in fact implies (9.3).

For the special case $p = 0$, this last condition is not necessary, since $g(j_1, j_2)$ and $h(j_1, j_2)$ do not take any values, and we have from (9.3) merely

$$(\tfrac{3}{2})^2 > (q + \tfrac{1}{2})^2. \tag{9.42}$$

The only integral value of q within these limits is $q = 0$ for $p = 0$.

IB. Suppose q is an integer, which, on account of the interchangeability of q and $-q - 1$ and of p and q may be taken to be such that $q \nless p \nless 0$.

For $q = p$, the argument of (IA) applies, and the only possible values are $p = 0$, $q = 0$, which can be regarded as a special case of (IA).

For $q > p \geq 0$, since $j_1 + j_2$ can take only values which make the radicals in $f(j_1, j_2)$ and $k(j_1, j_2)$ real, $j_1 + j_2$ must take the values $q + 1, q + 2, \cdots$, in $f(j_1, j_2)$ and $q, q + 1, \cdots$, in $k(j_1, j_2)$, and so $q, q + 1, \cdots$, in $g(j_1, j_2)$ and $h(j_1, j_2)$; and since $j_1 - j_2$ can only take values which make the radicals in $g(j_1, j_2)$ and $h(j_1, j_2)$ real, $j_1 - j_2$ must take: either the values $-q, -q + 1, \cdots$, $-p - 2$, in $g(j_1, j_2)$, $-q + 1, -q + 2, \cdots, -p - 1$, in $h(j_1, j_2)$, and $-q$, $-q + 1, \cdots, -p - 1$, in $f(j_1, j_2)$ and $k(j_1, j_2)$; or the values $p + 1, p + 2, \cdots$, $q - 1$, in $g(j_1, j_2)$, $p + 2, p + 3, \cdots, q$, in $h(j_1, j_2)$, and $p + 1, p + 2, \cdots, q$, in $f(j_1, j_2)$ and $k(j_1, j_2)$: these results hold even for $q - 1 = p > 0$, when there are no values for $g(j_1, j_2)$ and $h(j_1, j_2)$.

For $q > p = 0$, the above two representations still exist with the expected limiting forms when $q - 1 = p = 0$: the limiting cases of equality in (9.41) above are reducible to the sum of these two representations. In addition a solution is possible with $j_1 - j_2 = 0$ in $f(j_1, j_2)$ and $k(j_1, j_2)$ and no values for

$g(j_1, j_2)$ and $h(j_1, j_2)$; then $j_1 + j_2$ takes the values $q + 1, q + 2, \cdots$, in $f(j_1, j_2)$ and $q, q + 1, \cdots$, in $g(j_1, j_2)$. For $p = 0$, $q = 1$, this gives the limiting case of equality in (9.42) above.

Rearranging these results to correspond to $(p + \frac{1}{2})^2 \nless (q + \frac{1}{2})^2$, and describing the representations by the values which $j_1 + j_2$ and $j_1 - j_2$ take in β, β', we have:—

For

$$-\infty < (q + \tfrac{1}{2})^2 < (\tfrac{1}{2})^2,$$
$$(p + \tfrac{1}{2})^2 = (\tfrac{1}{2})^2, (\tfrac{3}{2})^2, (\tfrac{5}{2})^2, \cdots,$$

one representation each with

$$j_1 + j_2 = p, p + 1, p + 2, \cdots,$$
$$j_1 - j_2 = -p, -p + 1, \cdots, p.$$

For

$$(q + \tfrac{1}{2})^2 = (\tfrac{1}{2})^2,$$
$$(\tfrac{1}{2})^2 \ngtr (p + \tfrac{1}{2})^2 < (\tfrac{3}{2})^2,$$

one representation each with

$$j_1 + j_2 = 0, 1, 2, \cdots,$$
$$j_1 - j_2 = 0.$$

For

$$(q + \tfrac{1}{2})^2 = (\tfrac{1}{2})^2,$$
$$(p + \tfrac{1}{2})^2 = (\tfrac{3}{2})^2, (\tfrac{5}{2})^2, (\tfrac{7}{2})^2, \cdots,$$

three representations each with

$$j_1 + j_2 = p, p + 1, p + 2, \cdots,$$

and with

$$j_1 - j_2 = -p, -p + 1, \cdots, -1,$$

or

$$j_1 - j_2 = 0,$$

or

$$j_1 - j_2 = 1, 2, \cdots, p.$$

For

$$(q + \tfrac{1}{2})^2 = (\tfrac{3}{2})^2, (\tfrac{5}{2})^2, (\tfrac{7}{2})^2, \cdots,$$
$$(p + \tfrac{1}{2})^2 = (\tfrac{5}{2})^2, (\tfrac{7}{2})^2, (\tfrac{9}{2})^2, \cdots, > (q + \tfrac{1}{2})^2,$$

two representations each with

$$j_1 + j_2 = p, p + 1, p + 2, \cdots,$$

and with

$$j_1 - j_2 = -p, -p + 1, \cdots, -q - 1,$$

or

$$j_1 - j_2 = q + 1, q + 2, \cdots, p.$$

II. When only half odd integral values of $j_1 + j_2$ and therefore also of $j_1 - j_2$ are involved, the same arguments as in (I) apply, but there are no special cases, and we have:—

For

$$-\infty < (q + \tfrac{1}{2})^2 \not> 0,$$

$$(p + \tfrac{1}{2})^2 = 1^2, 2^2, 3^2, \cdots,$$

one representation each with

$$j_1 + j_2 = p, p + 1, p + 2, \cdots,$$

$$j_1 - j_2 = -p, -p + 1, \cdots, p.$$

For

$$(q + \tfrac{1}{2})^2 = 1^2, 2^2, 3^2, \cdots,$$

$$(p + \tfrac{1}{2})^2 = 2^2, 3^2, 4^2, \cdots, > (q + \tfrac{1}{2})^2,$$

two representations each with

$$j_1 + j_2 = p, p + 1, p + 2, \cdots$$

and with

$$j_1 - j_2 = -p, -p + 1, \cdots, -q - 1,$$

or

$$j_1 - j_2 = q + 1, q + 2, \cdots, p.$$

None of these representations are matrices of finite order. If it is not required that the representation be Hermitian, there are many further possibilities, including, both for integral values of p and q and for half odd integral values, finite sets where $j_1 + j_2$ takes the values $q, q + 1, \cdots, p - 1$, and $j_1 - j_2$ takes the values $-q, -q + 1, \cdots, q$; which are indeed complex transformations of the Hermitian representations of the operators of the rotation group in five dimensions, from which the algebraic form of (7.41) and (7.42) could have been derived.

Case (I) above corresponds to one-valued, case (II) to two-valued representations in the De Sitter space.

Ohio State University
Columbus, Ohio.

Annals of Mathematics
Vol. 42, No. 1, January, 1941

DER DREIERSTOSS

Von Carl Ludwig Siegel

(Received June 27, 1940)

Einleitung

Es seien A_1, A_2, A_3 drei Massenpunkte im Raum, die sich nach dem Newtonschen Gravitationsgesetz anziehen. Es seien x_k, y_k, z_k die rechtwinkligen kartesischen Koordinaten des Punktes A_k ($k = 1, 2, 3$) und m_k seine Masse; ferner seien die Abstände A_2A_3, A_3A_1, A_1A_2 mit r_1, r_2, r_3 bezeichnet. Setzt man

$$U = \frac{m_2 m_3}{r_1} + \frac{m_3 m_1}{r_2} + \frac{m_1 m_2}{r_3} \tag{1}$$

und versteht unter w_k eine der drei Koordinaten x_k, y_k, z_k, so lauten die Differentialgleichungen des Dreikörperproblems

$$m_k \frac{d^2 w_k}{dt^2} = \frac{\partial U}{\partial w_k} \qquad (w_k = x_k, y_k, z_k; k = 1, 2, 3). \tag{2}$$

Man gebe zur Zeit $t = t_0$ irgend welche endlichen Anfangswerte der 9 Koordinaten w_k, für welche die Abstände r_1, r_2, r_3 sämtlich grösser als 0 sind, und ausserdem beliebige endliche Anfangswerte der 9 Geschwindigkeitskomponenten $\frac{dw_k}{dt}$. Nach einem bekannten Existenzsatz aus der Theorie der Differentialgleichungen sind dann die durch jene Anfangswerte bestimmten Lösungen des Systemes (2) in der Umgebung von $t = t_0$ reguläre analytische Funktionen von t. Wir denken uns die Lösungen längs der reellen Achse von $t = t_0$ aus nach beiden Seiten analytisch fortgesetzt und nehmen an, dass wir auf diese Weise zu einem singulären Punkt $t = t_1$ einer der Funktionen w_k gelangen. Da die Differentialgleichungen (2) in sich übergehen, wenn zu t eine beliebige Konstante addiert wird und wenn t durch $-t$ ersetzt wird, so kann man $t_1 = 0$, $t_0 > 0$ voraussetzen.

Wie Sundman[1] gezeigt hat, bestehen bei dem Grenzübergang $t \to 0$, $t > 0$ nur die folgenden beiden Möglichkeiten: Entweder strebt genau eine der drei positiven Zahlen r_1, r_2, r_3 gegen 0, während die beiden andern einen positiven Grenzwert haben, oder aber sie streben alle drei gegen 0. Wir wollen diese beiden Fälle weiterhin als Zweierstoss und Dreierstoss bezeichnen. Der Zweierstoss wurde zuerst für das restringierte Dreikörperproblem von Levi-Civitá[2] und

[1] K. F. Sundman, *Recherches sur le problème des trois corps*, Acta Societatis Scientiarum Fennicae, Bd. 34 (1907), Nr. 6.

[2] T. Levi-Civitá, *Traiettorie singolari ed urti nel problema ristretto dei tre corpi*, Annali di matematica pura ed applicata, Ser. III^a, Bd. 9 (1904), S. 1–32.

dann allgemein von Bisconcini[3] behandelt. Sundman führte diese Untersuchungen weiter und zeigte insbesondere, dass beim Zweierstoss der Punkt $t = 0$ ein algebraischer Verzweigungspunkt zweiter Ordnung für die Lösungen ist und dass die Koordinaten w_k in der Umgebung dieses Punktes reguläre Funktionen der Ortsuniformisierenden $t^{\frac{1}{2}}$ sind.

Auch der Dreierstoss wurde von Sundman eingehend untersucht. Er bewies, dass beim Dreierstoss die Ausdrücke $r_k t^{-2/3}$ $(k = 1, 2, 3)$ für $t \to 0$ positive Grenzwerte $\hat{r}_k$ haben, für welche entweder $\hat{r}_1 = \hat{r}_2 = \hat{r}_3$ oder $\hat{r}_2 = \hat{r}_3 + \hat{r}_1$ oder $\hat{r}_3 = \hat{r}_1 + \hat{r}_2$ oder $\hat{r}_1 = \hat{r}_2 + \hat{r}_3$ gilt. Da die drei letzten dieser 4 Fälle durch zyklische Vertauschung der Indizes 1, 2, 3 ineinander übergeführt werden können, so wollen wir uns im folgenden nur noch mit dem ersten und zweiten beschäftigen. Zeichnet man ein Dreieck aus den Seiten $\hat{r}_1$, $\hat{r}_2$, $\hat{r}_3$, so erhält man im ersten Fall ein gleichseitiges Dreieck, im zweiten Fall drei Punkte einer Geraden; wir wollen daher von dem gleichseitigen und dem geradlinigen Fall sprechen. Ferner zeigte Sundman, dass bei jeder Dreierstossbahn die Ebene des Dreiecks $A_1A_2A_3$ für variables t nur eine Parallelverschiebung mit konstanter Geschwindigkeit erleidet. Da bekanntlich die Differentialgleichungen des Dreikörperproblems ungeändert bleiben, wenn man ein beliebiges neues rechtwinkliges Koordinatensystem einführt, das einer Translation mit konstanter Geschwindigkeit oder einer Drehung um konstante Winkel unterworfen wird, so braucht man für die weitere Untersuchung des Dreierstosses nur noch das ebene Dreikörperproblem zu behandeln und kann $z_k = 0$ $(k = 1, 2, 3)$ voraussetzen.

Aus den Resultaten Sundmans folgt sofort, dass beim Dreierstoss die drei Winkel des Dreiecks $A_1A_2A_3$ für $t \to 0$ Grenzwerte haben; im gleichseitigen Fall streben sie nämlich sämtlich gegen $\frac{1}{3}\pi$ und im geradlinigen Fall strebt der Winkel bei A_2 gegen π, während die beiden anderen Winkel gegen 0 streben. Es war aber bisher nicht bekannt, ob auch die Winkel der Dreiecksseiten mit den Koordinatenachsen für $t \to 0$ Grenzwerte haben, d. h. ob die drei Körper in bestimmten Richtungen zusammenstossen. Diese Frage wird im folgenden bejahend beantwortet werden. In engem Zusammenhang mit der Entscheidung dieser Frage steht das Problem der Entwickelbarkeit der Koordinaten der kollidierenden Massenpunkte in irreguläre Potenzreihen der Variabeln t. Wir werden solche Reihenentwicklungen wirklich aufstellen und damit zugleich beweisen, dass beim Dreierstoss der Punkt $t = 0$ im allgemeinen ein logarithmischer Verzweigungspunkt für die Lösungen ist. Dieses Ergebnis bezeichnet einen wesentlichen Unterschied gegenüber dem Zweierstoss. Die Reihenentwicklungen werden uns endlich einen vollen Überblick über sämtliche Dreierstossbahnen in der Nähe von $t = 0$ verschaffen. Sieht man zwei Lösungen nicht als verschieden an, wenn sie durch eine Drehung des Koordinatensystems um konstante Winkel oder durch eine Translation mit konstanter Geschwindigkeit ineinander übergeführt werden können, so zeigt es sich, dass in der Nähe

[3] G. Bisconcini, *Sur le problème des trois corps*, Acta Mathematica, Bd. 30 (1906), S. 49–92.

von $t = 0$ die sämtlichen Dreierstossbahnen im gleichseitigen Fall von 3 Parametern analytisch abhängen und im geradlinigen Fall von 2 Parametern.

Wir wollen das weiterhin abzuleitende hauptsächliche Resultat unserer Untersuchung noch präzis formulieren. Zur Abkürzung werde gesetzt

$$a = \frac{m_2 m_3 + m_3 m_1 + m_1 m_2}{(m_1 + m_2 + m_3)^2}$$

und

$$b = \frac{m_1\{1 + (1 - \omega)^{-1} + (1 - \omega)^{-2}\} + m_3(1 + \omega^{-1} + \omega^{-2})}{m_1 + m_2\{\omega^{-2} + (1 - \omega)^{-2}\} + m_3},$$

wo ω die im Intervall $0 < \omega < 1$ gelegene Lösung der Gleichung

$$m_1\{(1 - \omega)^{-2} - (1 - \omega)\} + m_2\{\omega(1 - \omega)^{-2} - (1 - \omega)\omega^{-2}\} + m_3(\omega - \omega^{-2}) = 0$$

bedeutet; ferner sei

$$a_1 = \tfrac{1}{6}\{-1 + [13 + 12\,(1 - 3a)^{\frac{1}{2}}]^{\frac{1}{2}}\},\ a_2 = \tfrac{1}{6}\{-1 + [13 - 12\,(1 - 3a)^{\frac{1}{2}}]^{\frac{1}{2}}\},$$

$$b_1 = \tfrac{1}{6}\{-1 + (25 + 16b)^{\frac{1}{2}}\}.$$

Ist dann weder $\dfrac{2}{3a_2}$ noch $\dfrac{a_1}{a_2}$ eine ganze Zahl, so lassen sich sämtliche Koordinaten x_k, y_k $(k = 1, 2, 3)$ im gleichseitigen Falle des Dreierstosses in der Form

$$x_k = t^{2/3} x_k^*(u_1, u_2, u_3), \qquad y_k = t^{2/3} y_k^*(u_1, u_2, u_3) \qquad (k = 1, 2, 3)$$

ausdrücken, wo $x_k^*(u_1, u_2, u_3)$ und $y_k^*(u_1, u_2, u_3)$ Potenzreihen in den Grössen

$$u_1 = \alpha_1 t^{2/3}, \qquad u_2 = \alpha_2 t^{a_1}, \qquad u_3 = \alpha_3 t^{a_2}$$

mit konstanten α_1, α_2, α_3 bedeuten; dabei hängen die Koeffizienten dieser Potenzreihen nur von m_1, m_2, m_3 ab. Die Werte α_1, α_2, α_3 sind eindeutig durch die Dreierstossbahn bestimmt, und umgekehrt liefert jedes System von reellen Werten α_1, α_2, α_3 wieder eine Dreierstossbahn, für welche der gleichseitige Fall vorliegt. Ist ferner $\dfrac{3b_1}{2}$ keine ganze Zahl, so haben die Koordinaten im geradlinigen Fall des Dreierstosses die Form

$$x_k = t^{2/3} x_k^*(v_1, v_2), \qquad y_k = t^{2/3} y_k^*(v_1 . v_2) \qquad (k = 1, 2, 3),$$

wo $x_k^*(v_1, v_2)$ und $y_k^*(v_1, v_2)$ Potenzreihen in

$$v_1 = \beta_1 t^{2/3}, \qquad v_2 = \beta_2 t^{b_1}$$

mit konstanten durch die Bahn eindeutig bestimmten Werten β_1, β_2 bedeuten; umgekehrt ergibt jedes reelle System β_1, β_2 eine Dreierstossbahn für den geradlinigen Fall. Ist $\dfrac{2}{3a_2}$ eine ganze Zahl g, so bleiben die obigen Aussagen bestehen, wenn man darin u_1 durch die Formel

$$u_1 = t^{2/3}(\alpha_1 + c_1 \alpha_2^g \log t)$$

erklärt, wo c_1 eine gewisse nur von m_1, m_2, m_3 abhängige Konstante bedeutet. Im Falle eines ganzzahligen $\frac{a_1}{a_2} = h$ hat man entsprechend

$$u_2 = t^{a_1}(\alpha_2 + c_2\alpha_3^h \log t)$$

zu erklären, und für ganzzahliges $\frac{3b_1}{2} = j$ ist

$$v_2 = t^{b_1}(\beta_2 + c_3\beta_1^j \log t)$$

zu setzen, wobei c_2 und c_3 wieder nur von m_1, m_2, m_3 abhängen.

Es sei noch daran erinnert, dass wir im geradlinigen Falle zwei weitere von zwei Parametern abhängige Scharen von Dreierstossbahnen erhalten, wenn wir m_1, m_2, m_3 zyklisch vertauschen. Ausserdem kann man noch eine beliebige Drehung des Koordinatensystems um konstante Winkel vornehmen, sowie eine beliebige Translation mit konstanter Geschwindigkeit.

In den ersten 6 Paragraphen der vorliegenden Arbeit werden im wesentlichen die Ergebnisse, welche Sundman für den Dreierstoss erhalten hat, in etwas veränderter Art hergeleitet. In §7 werden spezielle Dreierstossbahnen, die in den bekannten Lagrangeschen partikulären Lösungen des Dreikörperproblems enthalten sind, kurz besprochen. Als weitere Vorbereitung wird in §8 unter Anwendung der Jacobi-Hamiltonschen Theorie die Transformation der Differentialgleichungen des Dreikörperproblems behandelt, welche man in der Astronomie als Elimination der Knoten bezeichnet.[4] Es wird dann die Bestimmung aller Dreierstossbahnen zurückgeführt auf die Lösung folgender Aufgabe:

Es sei

$$\frac{d\delta_k}{ds} = f_k \qquad (k = 1, \cdots, n) \tag{3}$$

ein System von n Differentialgleichungen erster Ordnung, dessen rechte Seiten Potenzreihen der unbekannten Funktionen $\delta_1, \cdots, \delta_n$ sind, aber nicht die unabhängige Variable s explizit enthalten. Die Reihen f_k mögen keine konstanten Glieder haben, und es sei a_{kl} der Koeffizient von δ_l $(l = 1, \cdots, n)$ in den linearen Gliedern von f_k $(k = 1, \cdots, n)$. Von den charakteristischen Wurzeln der Matrix (a_{kl}) mögen genau p einen negativen Realteil haben und keine den Realteil 0. Man gebe sämtliche Lösungen des Systemes (3) an, welche für $s \to \infty$ den Grenzwert 0 haben.

Diese Aufgabe ist unter noch allgemeineren Voraussetzungen über die Funktionen f_k von Bohl[5] bearbeitet worden. Er zeigte auf sehr geistreiche Weise, dass jene Lösungen von p Parametern abhängen. Die von Bohl verwendeten topologischen Hilfssätze beruhen aber wesentlich auf indirekten Schlüssen und können nicht ohne weiteres zu einer Konstruktion der Lösungen benutzt werden.

[4] Vergl. S. 339–341 in E. T. Whittaker, *A treatise on the analytical dynamics of particles and rigid bodies, with an introduction to the problem of three bodies*, Cambridge (1904).

[5] P. Bohl, *Sur certaines équations différentielles d'un type général utilisables en mécanique*, Bulletin de la société mathématique de France, Bd. 38 (1910), S. 5–138.

Deshalb wird im folgenden in den Paragraphen 11 und 12 das Problem auf eine andere Art behandelt werden, durch Benutzung gewisser Reihentransformationen, welche die gesuchten Lösungen in expliziter Gestalt ergeben. Die irregulären Potenzreihen, durch welche wir die Lösungen ausdrücken werden, treten bereits in Poincaré's Untersuchungen über asymptotische Bahnkurven auf; dort fehlt aber gerade der Nachweis, dass diese Reihen wirklich alle asymptotischen Bahnen darstellen. Die Art, in der wir die Cauchysche Majorantenmethode anwenden, weicht von der üblichen ein wenig ab und bietet vielleicht auch ein selbständiges Interesse, da sie fast ohne Rechnung zum Ziele führt.

Im letzten Paragraphen werden schliesslich unsere Resultate über das System (3) angewendet auf die geeignet transformierten Differentialgleichungen des Dreikörperproblemes und ergeben den oben ausgesprochenen Satz über die Mannigfaltigkeit der Dreierstossbahnen.

1. Algebraische Vorbereitungen

Ist F eine Funktion der Zeit t, so soll die erste und die zweite Ableitung von F nach t in üblicher Weise mit $\dot{F}$ und $\ddot{F}$ bezeichnet werden. Unter $\gamma_1, \cdots, \gamma_{17}$ wollen wir weiterhin Grössen verstehen, die längs der zu betrachtenden Bahnkurve des Dreikörperproblems konstant sind. Zur Abkürzung werde noch

$$m = m_1 + m_2 + m_3 \tag{4}$$

gesetzt.

Wir schreiben zunächst die bekannten algebraischen Integrale des Dreikörperproblems auf, nämlich die Flächenintegrale, die Schwerpunktsintegrale und das Energieintegral. Die Flächenintegrale lauten

$$\begin{gathered} \sum_{k=1}^{3} m_k(y_k \dot{z}_k - z_k \dot{y}_k) = \gamma_1, \qquad \sum_{k=1}^{3} m_k(z_k \dot{x}_k - x_k \dot{z}_k) = \gamma_2, \\ \sum_{k=1}^{3} m_k(x_k \dot{y}_k - y_k \dot{x}_k) = \gamma_3. \end{gathered} \tag{5}$$

Die Schwerpunktsintegrale besagen, dass der Schwerpunkt der drei Massenpunkte sich geradlinig und gleichförmig bewegt, nämlich

$$\begin{gathered} \frac{1}{m} \sum_{k=1}^{3} m_k x_k = \gamma_4 t + \gamma_5, \qquad \frac{1}{m} \sum_{k=1}^{3} m_k y_k = \gamma_6 t + \gamma_7, \\ \frac{1}{m} \sum_{k=1}^{3} m_k z_k = \gamma_8 t + \gamma_9. \end{gathered}$$

Da die Differentialgleichungen (2) sich nicht ändern, wenn x_k, y_k, z_k durch $x_k + \gamma_4 t + \gamma_5, y_k + \gamma_6 t + \gamma_7, z_k + \gamma_8 t + \gamma_9$ $(k = 1, 2, 3)$ ersetzt werden, so kann man weiterhin voraussetzen, dass der Schwerpunkt fest im Koordinatenanfangspunkt liegt, dass also die Gleichungen

$$\sum_{k=1}^{3} m_k x_k = 0, \qquad \sum_{k=1}^{3} m_k y_k = 0, \qquad \sum_{k=1}^{3} m_k z_k = 0 \tag{6}$$

gelten. Bedeutet

$$T = \tfrac{1}{2} \sum_{k=1}^{3} m_k(\dot{x}_k^2 + \dot{y}_k^2 + \dot{z}_k^2) \tag{7}$$

die lebendige Kraft des Punktsystems, so ist

$$T - U = \gamma_{10} \tag{8}$$

das Energieintegral.

Wir setzen nun

$$J = \sum_{k=1}^{3} m_k(x_k^2 + y_k^2 + z_k^2). \tag{9}$$

Dann ist

$$\tfrac{1}{2}\dot{J} = \sum_{k=1}^{3} m_k(x_k\dot{x}_k + y_k\dot{y}_k + z_k\dot{z}_k), \tag{10}$$

$$\tfrac{1}{2}\ddot{J} = \sum_{k=1}^{3} m_k(\dot{x}_k^2 + \dot{y}_k^2 + \dot{z}_k^2) + \sum_{k=1}^{3} m_k(x_k\ddot{x}_k + y_k\ddot{y}_k + z_k\ddot{z}_k),$$

also nach (2) und (7)

$$\tfrac{1}{2}\ddot{J} = 2T + \sum_{k=1}^{3}\left(x_k \frac{\partial U}{\partial x_k} + y_k \frac{\partial U}{\partial y_k} + z_k \frac{\partial U}{\partial z_k}\right). \tag{11}$$

Da nun U eine homogene Funktion der Dimension -1 in den 9 Variabeln $x_1, \cdots, z_3$ ist, so ist nach einem bekannten Eulerschen Satze

$$\sum_{k=1}^{3}\left(x_k \frac{\partial U}{\partial x_k} + y_k \frac{\partial U}{\partial y_k} + z_k \frac{\partial U}{\partial z_k}\right) = -U$$

und (11) ergibt die Lagrangesche Formel

$$\tfrac{1}{2}\ddot{J} = 2T - U,$$

also nach (8)

$$\tfrac{1}{2}\ddot{J} = T + \gamma_{10}, \tag{12}$$

$$\tfrac{1}{2}\ddot{J} = U + 2\gamma_{10}. \tag{13}$$

Sind $B_1, \cdots, B_n$ und $C_1, \cdots, C_n$ zwei Reihen von je n Grössen und bezeichnet man mit

$$D_{pq} = B_pC_q - B_qC_p \qquad (1 \leqq p < q \leqq n)$$

die $n(n-1)/2$ zweireihigen Unterdeterminanten der aus diesen beiden Reihen gebildeten Matrix, so gilt identisch

$$\sum_{p=1}^{n} B_p^2 \sum_{p=1}^{n} C_p^2 - \left(\sum_{p=1}^{n} B_pC_p\right)^2 = \sum_{p,q} D_{pq}^2. \tag{14}$$

Wir wählen $n = 9$ und identifizieren die Paare B_p, C_p $(p = 1, \cdots, 9)$ mit $w_k\sqrt{m_k}$, $\dot{w}_k\sqrt{m_k}$ $(w_k = x_k, y_k, z_k; k = 1, 2, 3)$. Behält man die Abkürzung

D_{pq} bei, so geht (14) zufolge (7), (9), (10) über in

$$2TJ - \tfrac{1}{4}\dot{J}^2 = \sum_{p,q} D_{pq}^2. \tag{15}$$

Also gilt insbesondere die Ungleichung

$$\dot{J}^2 \leqq 8JT. \tag{16}$$

2. Die Singularitäten der Bahnkurve

Wir wenden auf die Differentialgleichungen des Dreikörperproblems folgenden bekannten Existenzsatz an:

Es seien $F_1, \cdots, F_n$ Funktionen von n Variabeln $s_1, \cdots, s_n$, die in der Umgebung eines Punktes $s_1 = \sigma_1, \cdots, s_n = \sigma_n$ in Reihen nach Potenzen von $s_1 - \sigma_1, \cdots, s_n - \sigma_n$ entwickelbar sind. Es sei δ eine positive Zahl, für welche die Potenzreihen in dem Gebiete

$$| s_1 - \sigma_1 | \leqq \delta, \cdots, | s_n - \sigma_n | \leqq \delta \tag{17}$$

konvergieren; ferner sei K eine gemeinsame obere Schranke der Werte $| F_1 |, \cdots, | F_n |$ in diesem Gebiete. Es sei t eine reelle Variable und t_0 irgend ein Wert von t. Es gibt genau ein System von n Funktionen $\varphi_1, \cdots, \varphi_n$ der reellen Variabeln t mit folgenden Eigenschaften:

1) Die Funktion $\varphi_k(t)$ ist differentiierbar in einer Umgebung von $t = t_0$ und es ist $\varphi_k(t_0) = \sigma_k$ $(k = 1, \cdots, n)$;

2) in dieser Umgebung von $t = t_0$ genügen $s_1 = \varphi_1(t), \cdots, s_n = \varphi_n(t)$ den Differentialgleichungen

$$\frac{ds_k}{dt} = F_k \qquad (k = 1, \cdots, n).$$

Es gibt ausserdem eine nur von δ, K und n abhängige positive Zahl τ, sodass die Funktionen $s_k = \varphi_k(t)$ in Reihen nach Potenzen von $t - t_0$ entwickelbar sind, welche sämtlich für $| t - t_0 | < \tau$ konvergieren und in diesem Gebiete den Ungleichungen (17) genügen.

Um diesen Satz auf das System (2) anzuwenden, wählen wir $n = 18$ und setzen für $k = 1, 2, 3$

$$x_k = s_k, \qquad y_k = s_{k+3}, \qquad z_k = s_{k+6},$$

$$\dot{x}_k = s_{k+9} = F_k, \qquad \dot{y}_k = s_{k+12} = F_{k+3}, \qquad \dot{z}_k = s_{k+15} = F_{k+6},$$

$$\frac{1}{m_k}\frac{\partial U}{\partial x_k} = F_{k+9}, \qquad \frac{1}{m_k}\frac{\partial U}{\partial y_k} = F_{k+12}, \qquad \frac{1}{m_k}\frac{\partial U}{\partial z_k} = F_{k+15}.$$

Zur Zeit $t = t_0$ seien irgend welche endlichen reellen Anfangswerte $\sigma_1, \cdots, \sigma_{18}$ der 18 Grössen $x_1, \cdots, \dot{z}_3$ gegeben, für welche keine Kollision vorliegt, also die Abstände r_1, r_2, r_3 grösser als 0 sind. Wir berechnen aus den Zahlen $\sigma_1, \cdots, \sigma_{18}$ nach (1) und (8) den Anfangswert $U = U_0$ und den Wert der Konstanten γ_{10} des Energieintegrales. Ist dann K_1 irgend eine obere Schranke für

U_0, so gelten nach (1), (7), (8) für die Anfangswerte die Ungleichungen

$$r_1^{-1} < \frac{K_1}{m_2 m_3}, \qquad r_2^{-1} < \frac{K_1}{m_3 m_1}, \qquad r_3^{-1} < \frac{K_1}{m_1 m_2},$$

$$\dot{x}_k^2 + \dot{y}_k^2 + \dot{z}_k^2 \leqq 2(K_1 + \gamma_{10}) m_k^{-1} \qquad (k = 1, 2, 3).$$

Folglich lassen sich zwei positive Grössen δ und K als Funktionen von m_1, m_2, m_3, K_1 und γ_{10} allein so wählen, dass die Voraussetzungen des Satzes erfüllt sind. Die durch die Anfangswerte eindeutig bestimmte Lösung $s_k = \varphi_k(t)$ $(k = 1, \cdots, 18)$ der Differentialgleichungen ist dann regulär in einem Zeitintervall $t_0 - \tau < t < t_0 + \tau$, wo die positive Zahl τ nur von m_1, m_2, m_3, K_1, γ_{10} abhängt.

Wir betrachten jetzt, von $t = t_0$ ausgehend, die Funktionen $\varphi_k(t)$ für fallende reelle Werte von t. Entweder sind sie sämtlich regulär für alle endlichen reellen $t \leqq t_0$ oder aber es gibt einen endlichen Wert $t = t_1 < t_0$, sodass alle Funktionen im links offenen Intervall $t_1 < t \leqq t_0$ regulär sind und mindestens eine von ihnen für $t = t_1$ singulär ist. Eine analoge Aussage gilt für $t \geqq t_0$ und wachsende Werte von t. Da die Differentialgleichungen (2) bei der Transformation $t \to -t$ in sich übergehen, so kann man sich auf die Untersuchung für fallendes $t \leqq t_0$ beschränken. Weil die Differentialgleichungen die Zeit t nicht explizit enthalten, so kann man ausserdem noch $t_1 = 0$ voraussetzen. Wegen der Regularität für $t_0 - \tau < t < t_0 + \tau$ ist dann $\tau \leqq t_0$. Nun sei $0 < t_2 < \tau$ und $U = U_2$ der Wert von U für $t = t_2$. Wäre auch $U_2 < K_1$, so wären die Funktionen $\varphi_k(t)$ auch sämtlich in dem Intervall $t_2 - \tau < t < t_2 + \tau$ regulär; aber dies ist ein Widerspruch, da das Intervall den singulären Punkt $t = 0$ enthält. Für $0 < t < \tau$ gilt daher $U \geqq K_1$, und dabei hängt τ nur von m_1, m_2, m_3, γ_{10} und K_1 ab. Da K_1 beliebig gross sein kann, so folgt, dass U über alle Schranken wächst, wenn t zu 0 abnimmt. Dies bedeutet, dass für $t \to 0$ der kleinste der 3 Abstände r_1, r_2, r_3 gegen 0 strebt.

In folgenden bedeuten τ_1, τ_2, τ_3 geeignete hinreichend klein zu wählende positive Zahlen, die nur von den Massen m_1, m_2, m_3 und den gegebenen Anfangswerten von $x_1, \cdots, \dot{z}_3$ abhängen. Auf der betrachteten Bahnkurve ist $U \to \infty$ für $t \to 0$, also nach (13)

$$\ddot{J} > 0 \qquad (0 < t \leqq \tau_1). \tag{18}$$

Daher ist J im Intervall $0 < t \leqq \tau_1$ eine konvexe Funktion von t und hat folglich für $t \to 0$ einen Grenzwert J_0, der positiv oder 0 sein kann. Im Falle $J_0 > 0$ folgt aus (6) und (9), dass der grösste der drei Abstände r_1, r_2, r_3 für $t \to 0$ oberhalb einer positiven Schranke bleibt. Da andererseits der kleinste dieser Abstände gegen 0 strebt und sie sämtlich für $t > 0$ stetige Funktionen von t sind, so ergibt sich, dass für $t \to 0$ eine bestimmte der Dreiecksseiten r_k gegen 0 strebt, während die beiden anderen oberhalb einer positiven Schranke bleiben; es stossen dann also zur Zeit $t = 0$ genau zwei von den Körpern zusammen. Dieser Zweierstoss ist von Sundman vollständig untersucht worden. Er hat gezeigt,

dass die Koordinaten der drei Körper in der Umgebung von $t = 0$ reguläre Funktionen der Ortsuniformisierenden $t^{1/3}$ sind.

Weiterhin beschäftigen wir uns dauernd mit dem Fall $J_0 = 0$. Dann streben aber nach (9) für $t \to 0$ alle drei Abstände r_1, r_2, r_3 gegen 0 und es stossen zur Zeit $t = 0$ alle drei Körper im Nullpunkt zusammen. Es soll genauer untersucht werden, in welcher Weise dieser Dreierstoss vor sich geht.

3. Das asymptotische Verhalten von J und $\dot{J}$

Nach (18) ist die positive Funktion J konvex im Intervall $0 < t \leqq \tau_1$, ferner strebt J gegen 0 für $t \to 0$. Folglich gilt

$$\dot{J} > 0 \qquad (0 < t \leqq \tau_1). \tag{19}$$

Andererseits ist nach (12)

$$\ddot{J}J^{-1/4} - \tfrac{1}{4}\dot{J}^2 J^{-5/4} = \tfrac{1}{4}(8JT - \dot{J}^2)J^{-5/4} + 2\gamma_{10}J^{-1/4},$$

und hierin ist die linke Seite gerade die Ableitung der Funktion $\dot{J}J^{-1/4}$ nach t. Integriert man diese Gleichung zwischen den Grenzen t und τ_1 und bezeichnet mit J_1 und $\dot{J}_1$ die Werte von J und $\dot{J}$ für $t = \tau_1$, so folgt

$$\dot{J}_1 J_1^{-1/4} - \dot{J}J^{-1/4} = \tfrac{1}{4}\int^{\tau_1} (8JT - \dot{J}^2)J^{-5/4}\,dt + 2\gamma_{10}\int_t^{\tau_1} J^{-1/4}\,dt. \tag{20}$$

Wir wollen nun beweisen, dass die beiden Integrale auf der rechten Seite für $t \to 0$ endliche Grenzwerte haben. Wir verstehen weiterhin unter $\mu_1, \cdots, \mu_5$ gewisse positive Zahlen, die genügend klein zu wählen sind und nur von m_1, m_2, m_3 abhängen. Da der Nullpunkt im Schwerpunkt des Dreiecks $A_1A_2A_3$ liegt, so ist

$$J > \mu_1(r_1^2 + r_2^2 + r_3^2)$$

und folglich

$$U > \mu_2 J^{-\frac{1}{2}},$$

also auch

$$U + 2\gamma_{10} > \mu_3 J^{-\frac{1}{2}} \qquad (0 < t < \tau_2).$$

Im Intervall $0 < t < \tau_2$ gilt dann nach (13) und (19)

$$(\dot{J}^2)^{\cdot} = 2\dot{J}\ddot{J} > 4\mu_3\dot{J}J^{-\frac{1}{2}}$$

$$\dot{J}^2 > 8\mu_3 J^{\frac{1}{2}}$$

$$\dot{J}J^{-1/4} > \mu_4 \tag{21}$$

$$J > \mu_5 t^{4/3}.$$

Demnach konvergiert das zweite Integral in (20) bis nach $t = 0$. Zufolge (19) und (20) ist dann das erste Integral in (20) für $t \to 0$ nach oben beschränkt; da

aber sein Integrand wegen (16) nicht-negativ ist, so konvergiert es ebenfalls bis nach $t = 0$. Nach (20) existiert also

$$\lim_{t \to 0} \dot{J} J^{-\frac{1}{4}} = \gamma_{11}, \tag{22}$$

und zwar ist dieser Grenzwert zufolge (21) eine positive Zahl. Durch Integration von $\dot{J} J^{-1/4}$ folgt aus (22) die asymptotische Gleichung

$$J^{3/4} \sim \tfrac{3}{4}\gamma_{11} t \qquad (t \to 0)$$

$$J \sim \lambda t^{4/3} \tag{23}$$

mit

$$\lambda = (\tfrac{3}{4}\gamma_{11})^{4/3} > 0, \tag{24}$$

also nach (22)

$$\dot{J} \sim \tfrac{4}{3}\lambda t^{1/3}. \tag{25}$$

Damit ist das asymptotische Verhalten von J und $\dot{J}$ klargelegt.

4. Das asymptotische Verhalten von U und T

Setzt man

$$(8JT - \dot{J}^2)t^{-2/3} = g(t), \tag{26}$$

so ist nach (16)

$$g(t) \geqq 0 \tag{27}$$

und nach §3 das Integral

$$\int_0^{\tau_1} g(t) t^{2/3} J^{-5/4}\, dt$$

konvergent. Zufolge (23), (24), (25) konvergiert dann auch

$$\int_0^{\tau_1} g(t) \frac{dt}{t}. \tag{28}$$

Wir wollen beweisen, dass $g(t)$ für $t \to 0$ den Grenzwert 0 besitzt. Wegen (27) und der Konvergenz des Integrales (28) ist jedenfalls

$$\liminf_{t \to 0} g(t) = 0. \tag{29}$$

Wäre nun

$$\limsup_{t \to 0} g(t) > 0,$$

so könnte man wegen (29) und der Stetigkeit von $g(t)$ für $t > 0$ eine positive Zahl γ_{12} und eine monoton zu 0 abnehmende Folge $\tau_1 > t_1 > t_2 > \cdots$ so finden, dass

(30) $$g(t_{2n}) = \gamma_{12}\,, \qquad g(t_{2n-1}) = 3\gamma_{12} \qquad (n = 1, 2, 3, \cdots)$$

und in jedem Intervall $t_{2n} \leqq t \leqq t_{2n-1}$ die Ungleichung

(31) $$\gamma_{12} \leqq g(t) \leqq 3\gamma_{12}$$

gilt.

Nach (26) ist

$$T = \tfrac{1}{8}\{\dot{J}^2 + g(t)t^{2/3}\}J^{-1},$$

also nach (23), (25), (31)

(32) $$T < \gamma_{13}t^{-2/3} \qquad (t_{2n} \leqq t \leqq t_{2n-1}\,;\; n = 1, 2, 3, \cdots).$$

Ferner ist

$$\dot{T} = \dot{U} = \sum_{k=1}^{3}\left(\frac{\partial U}{\partial x_k}\dot{x}_k + \frac{\partial U}{\partial y_k}\dot{y}_k + \frac{\partial U}{\partial z_k}\dot{z}_k\right),$$

also nach (1), (7), (8), (32)

(33) $$|\,\dot{T}\,| < \gamma_{14}I^{-2}T^{\frac{1}{2}} < \gamma_{15}t^{-5/3},$$

wieder im Intervall $t_{2n} \leqq t \leqq t_{2n-1}$ $(n = 1, 2, 3, \cdots)$. Folglich gilt dort auch nach (23), (25), (32), (33) die Ungleichung

$$|\,(8JTt^{-2/3})^{\cdot}\,| < \gamma_{16}t^{-1},$$

und die Funktion $8JTt^{-2/3}$ ändert sich daher im Intervall $t_{2n} \leqq t \leqq t_{2n-1}$ um weniger als

$$\gamma_{16}\int_{t_{2n}}^{t_{2n-1}}\frac{dt}{t}.$$

In dem gleichen Intervall ändert sich ferner die Funktion $\dot{J}^2t^{-2/3}$ zufolge (25) um weniger als γ_{12}, wenn nur $t_{2n-1} < \tau_3$ ist, also für alle genügend grossen Werte von n. Nach (26) und (30) erhält man dann

$$2\gamma_{12} = g(t_{2n-1}) - g(t_{2n}) < \gamma_{12} + \gamma_{16}\int_{t_{2n}}^{t_{2n-1}}\frac{dt}{t},$$

also

$$\int_{t_{2n}}^{t_{2n-1}}\frac{dt}{t} > \frac{\gamma_{12}}{\gamma_{16}}$$

und nach (31)

(34) $$\int_{t_{2n}}^{t_{2n-1}} g(t)\,\frac{dt}{t} > \gamma_{17} > 0\,,$$

für alle genügend grossen n. Durch Summation über n folgt aber aus (34) ein Widerspruch gegen die Konvergenz des Integrales (28).

Damit ist bewiesen, dass $g(t)$ für $t \to 0$ den Grenzwert 0 besitzt. Es ist also auch

$$8JT - \dot{J}^2 = o(t^{2/3}) \tag{35}$$

und nach (23), (25)

$$T \sim \tfrac{2}{9}\lambda t^{-2/3}, \tag{36}$$

nach (8)

$$U \sim \tfrac{4}{9}\lambda t^{-2/3}. \tag{37}$$

5. Ebene Bewegung

Aus (15) und (35) ersicht man, dass jede der 36 in (15) auftretenden zweireihigen Determinanten D_{pq} für $t \to 0$ von kleinerer Grössenordnung als $t^{1/3}$ ist. Sind w und w_0 irgend zwei der 9 Koordinaten $x_1, \cdots, z_3$, so gilt also

$$w\dot{w}_0 - w_0\dot{w} = o(t^{1/3}). \tag{38}$$

Nach (5) haben daher die Flächenkonstanten $\gamma_1, \gamma_2, \gamma_3$ alle drei den Wert 0. Hieraus ergibt sich nun leicht, dass die Bewegung in einer festen Ebene durch den Nullpunkt vor sich geht:

Die Differentialgleichungen (2) sind invariant gegenüber einer Drehung des Koordinatensystems um konstante Winkel. Deshalb kann man annehmen, dass zur Zeit $t = t_0$ die drei Punkte A_1, A_2, A_3 in der Ebene $z = 0$ liegen. Wegen $\gamma_1 = 0, \gamma_2 = 0$ ist dann also

$$\sum_{k=1}^{3} m_k y_k \dot{z}_k = 0, \qquad \sum_{k=1}^{3} m_k x_k \dot{z}_k = 0 \tag{39}$$

für $t = t_0$. Nach (6) ist ferner

$$\sum_{k=1}^{3} m_k \dot{z}_k = 0. \tag{40}$$

Liegen nun die drei Massenpunkte zur Zeit $t = t_0$ nicht auf einer Geraden, so ist die aus den drei Zeilen x_1, x_2, x_3; y_1, y_2, y_3; 1, 1, 1 gebildete Determinante von 0 verschieden, und aus (39), (40) folgt das Verschwinden der drei Werte $\dot{z}_1, \dot{z}_2, \dot{z}_3$ für $t = t_0$. Liegen andererseits die drei Punkte zur Zeit $t = t_0$ auf einer Geraden, so kann man durch eine Drehung des Koordinatensystems erreichen, dass diese Gerade die x-Achse wird und ausserdem die Geschwindigkeitskomponente $\dot{z}_3$ für $t = t_0$ verschwindet. Nach (39) und (40) ist dann aber

$$m_1 x_1 \dot{z}_1 + m_2 x_2 \dot{z}_2 = 0, \qquad m_1 \dot{z}_1 + m_2 \dot{z}_2 = 0 \tag{41}$$

für $t = t_0$. Da zur Zeit $t = t_0$ keine Kollision stattfindet, so ist dann $x_1 \neq x_2$, und aus (41) folgt wieder das Verschwinden von $\dot{z}_1$ und $\dot{z}_2$ für $t = t_0$. In jedem Fall liegen also die Richtungen der Bewegungen der drei Massenpunkte zur Zeit $t = t_0$ ebenfalls in der Ebene $z = 0$, und aus den Differentialgleichungen

(2) folgt nach dem Eindeutigkeitssatz, dass die Bewegung ganz in der Ebene $z = 0$ erfolgt.

Wir können daher weiterhin $z_k = 0$ $(k = 1, 2, 3)$ annehmen und brauchen nur noch ebene Bewegungen zu betrachten.

6. Das asymptotische Verhalten der Dreiecksseiten

Bedeutet w eine beliebige der 6 Koordinaten x_k, y_k $(k = 1, 2, 3)$, so wollen wir die Abkürzung

$$w^* = wt^{-2/3} \tag{42}$$

einführen, also z.B. $x_1^* = x_1 t^{-2/3}$. Ist allgemeiner Φ eine homogene Funktion von $x_1, \cdots, y_3$, so wollen wir unter Φ^* den Wert verstehen, den man erhält, wenn man in Φ alle Variabeln w durch w^* ersetzt; z.B. ist also

$$r_k^* = r_k t^{-2/3}, \qquad U^* = Ut^{2/3}.$$

Nach (23) und (36) gilt die Abschätzung

$$w = O(t^{2/3}), \qquad \dot{w} = O(t^{-1/3}). \tag{43}$$

Hieraus folgt nach (42)

$$w^* = O(1), \qquad \dot{w}^* = \dot{w}t^{-2/3} - \tfrac{2}{3}wt^{-5/3} = O(t^{-1}). \tag{44}$$

Ferner ist nach (38), wenn w_0 ebenfalls eine Koordinate bedeutet,

$$w^*\dot{w}_0^* - w_0^*\dot{w}^* = wt^{-2/3}(\dot{w}_0 t^{-2/3} - \tfrac{2}{3}w_0 t^{-5/3}) - w_0 t^{-2/3}(\dot{w}t^{-2/3} - \tfrac{2}{3}wt^{-5/3}) = o(t^{-1}) \tag{45}$$

und nach (23), (26)

$$J^* \sim \lambda, \qquad \dot{J}^* = \dot{J}t^{-4/3} - \tfrac{4}{3}Jt^{-7/3} = o(t^{-1}). \tag{46}$$

Aus (44) und (45) erhält man

$$J^*\dot{w}^* - \tfrac{1}{2}w^*\dot{J}^* = \sum_{k=1}^{3} m_k\{x_k^*(x_k^*\dot{w}^* - w^*\dot{x}_k^*) + y_k^*(y_k^*\dot{w}^* - w^*\dot{y}_k^*)\} = o(t^{-1}) \tag{47}$$

und aus (44) und (46)

$$J^*\dot{w}^* - \tfrac{1}{2}w^*\dot{J}^* = \lambda\dot{w}^* + o(t^{-1}). \tag{48}$$

Aus (24), (47), (48) ergibt sich

$$\dot{w}^* = o(t^{-1}). \tag{49}$$

Es ist

$$\frac{\partial U^*}{\partial w^*} = \frac{\partial U}{\partial w} t^{4/3}$$

und folglich gehen die Differentialgleichungen (2) durch die Substitution (42)

über in

$$m_k(w_k^* t^{2/3})^{\cdot\cdot} = \frac{\partial U^*}{\partial w_k^*} t^{-4/3} \qquad (w_k = x_k, y_k; k = 1, 2, 3)$$

$$-\tfrac{2}{9} w_k^* + (\dot{w}_k^* t^{4/3})^{\cdot} t^{2/3} = \frac{1}{m_k} \frac{\partial U^*}{\partial w_k^*}. \tag{50}$$

Nun sei $\epsilon > 0$, $\epsilon \to 0$. Im Intervall $\epsilon \leqq t \leqq 2\epsilon$ gilt nach (49)

$$w^* = (w^*)_{t=\epsilon} + o\left(\int_\epsilon^{2\epsilon} \frac{dt}{t}\right) = (w^*)_{t=\epsilon} + o(1). \tag{51}$$

Ferner ist nach (37)

$$U^* \sim \tfrac{2}{9}\lambda;$$

also sind für $t \to 0$ die reziproken Werte der Grössen r_k^* $(k = 1, 2, 3)$ beschränkt und das gleiche gilt dann für die zweiten partiellen Ableitungen von U^* nach seinen Variabeln w^*. Nach dem Mittelwertsatz folgt aus (51) demnach

$$\frac{\partial U^*}{\partial w^*} = \left(\frac{\partial U^*}{\partial w^*}\right)_{t=\epsilon} + o(1) \qquad (\epsilon \leqq t \leqq 2\epsilon). \tag{52}$$

Ausserdem ist nach (49)

$$\int_\epsilon^t (\dot{w}^* t^{4/3})^{\cdot} t^{2/3}\, dt = [\dot{w}^* t^2]_\epsilon^t - \tfrac{2}{3} \int_\epsilon^t \dot{w}^* t\, dt = o(t) \qquad (\epsilon \leqq t \leqq 2\epsilon). \tag{53}$$

Wir integrieren jetzt die Gleichung (50) zwischen den Grenzen ϵ und 2ϵ und ersetzen nachträglich wieder ϵ durch t. Nach (51), (52), (53) folgt dann

$$-\tfrac{2}{9} w_k^* t + o(t) = \frac{1}{m_k} \frac{\partial U^*}{\partial w_k^*} t + o(t)$$

$$\tfrac{2}{9} w_k^* + \frac{1}{m_k} \frac{\partial U^*}{\partial w_k^*} \to 0 \qquad (t \to 0). \tag{54}$$

Diese Relation ist offenbar invariant bei beliebiger orthogonaler Transformation des Koordinatensystems. Wir wählen ein derartiges bewegliches Koordinatensystem durch den Schwerpunkt, dass die Strecke A_3A_1 parallel zur Abszissenachse ist und A_2 eine nicht-negative Ordinate hat. Sind X_k, Y_k die neuen Koordinaten von A_k $(k = 1, 2, 3)$, so ist also $X_1 > X_3$, $Y_1 = Y_3$, $Y_2 \geqq 0$. Setzt man noch

$$p_1 = X_1 - X_3, \qquad p_2 = X_2 - X_3, \qquad p_3 = Y_2 - Y_3, \tag{55}$$

so haben im neuen System die Punkte A_1 und A_2 in bezug auf A_3 die Relativkoordinaten p_1, 0 und p_2, p_3. Es sei wieder

$$X_k^* = X_k t^{-2/3}, \qquad Y_k^* = Y_k t^{-2/3}, \qquad p_k^* = p_k t^{-2/3}.$$

Benutzt man (54) mit Y_2^*, Y_3^* anstelle von w_k^*, so folgt durch Subtraktion

$$p_3^*\left(\frac{1}{r_2^{*3}} - \frac{1}{r_3^{*3}}\right) \to 0. \tag{56}$$

Verwendet man (54) analog für X_1^*, X_3^* und für X_2^*, X_3^*, so erhält man

$$\tfrac{2}{9}p_1^* + m_2\left(\frac{p_2^* - p_1^*}{r_3^{*3}} - \frac{p_2^*}{r_1^{*3}}\right) - (m_1 + m_3)\frac{p_1^*}{r_2^{*3}} \to 0 \tag{57}$$

$$\tfrac{2}{9}p_2^* + m_1\left(\frac{p_1^* - p_2^*}{r_3^{*3}} - \frac{p_1^*}{r_2^{*3}}\right) - (m_2 + m_3)\frac{p_2^*}{r_1^{*3}} \to 0. \tag{58}$$

Nach (23) sind die Werte p_k^* $(k = 1, 2, 3)$ beschränkt; nach (37) sind auch die reziproken Werte der r_k^* beschränkt. Wir betrachten jetzt irgend eine Folge $t \to 0$, für welche die zugehörigen Werte der p_k^* gegen Grenzwerte $\hat{p}_k$ streben. Es seien $\hat{r}_k$ $(k = 1, 2, 3)$ die Grenzwerte der Grössen r_k^*. Da der Schwerpunkt des Dreiecks im Nullpunkt liegt, so streben auch die Punkte (X_k^*, Y_k^*) gegen gewisse Grenzpunkte $\hat{A}_k$ $(k = 1, 2, 3)$. Zufolge (56) gilt

$$\hat{p}_3(\hat{r}_2^{-3} - \hat{r}_3^{-3}) = 0,$$

also entweder $\hat{p}_3 = 0$ oder $\hat{r}_2 = \hat{r}_3$.

Liegen die drei Punkte $\hat{A}_1$, $\hat{A}_2$, $\hat{A}_3$ nicht auf einer Geraden, so ist $\hat{p}_3 \neq 0$, also $\hat{r}_2 = \hat{r}_3$, und durch zyklische Vertauschung folgt auch $\hat{r}_3 = \hat{r}_1$; daher ist das Dreieck $\hat{A}_1\hat{A}_2\hat{A}_3$ dann gleichseitig. Bedeutet r die Dreiecksseite, so ist offenbar

$$\hat{p}_1 = r, \qquad \hat{p}_2 = \tfrac{1}{2}r, \qquad \hat{p}_3 = \tfrac{1}{2}r\sqrt{3}, \tag{59}$$

also nach (4) und (57)

$$\tfrac{2}{9}r^3 = m, \tag{60}$$

ferner nach (37)

$$\tfrac{2}{9}\lambda = (m_2m_3 + m_3m_1 + m_1m_2)r^{-1}. \tag{61}$$

Liegen $\hat{A}_1$, $\hat{A}_2$, $\hat{A}_3$ auf einer Geraden, so kann man nach etwaiger zyklischer Vertauschung der Indizes voraussetzen, dass $\hat{A}_2$ zwischen $\hat{A}_1$ und $\hat{A}_3$ gelegen ist. Setzt man

$$\hat{p}_1 = \rho, \qquad \hat{p}_2 = \omega\rho, \tag{62}$$

so ist also

$$0 < \omega < 1 \tag{63}$$

und aus (57), (58) folgen die Gleichungen

$$\tfrac{2}{9}\rho^3 = m_1 + m_3 + m_2\{\omega^{-2} + (1 - \omega)^{-2}\} \tag{64}$$

$$\tfrac{2}{9}\omega\rho^3 = m_1\{1 - (1 - \omega)^{-2}\} + (m_2 + m_3)\omega^{-2}. \tag{65}$$

Daher genügt ω der algebraischen Gleichung fünften Grades

(66) $m_1\{(1-\omega)^{-2}-(1-\omega)\}+m_2\{\omega(1-\omega)^{-2}-(1-\omega)\omega^{-2}\}+m_3(\omega-\omega^{-2})=0.$

Schreibt man sie in der Form

$$\frac{m_1+m_2\omega}{m_1+m_2\omega^{-2}}=\frac{m_3+m_2(1-\omega)}{m_3+m_2(1-\omega)^{-2}}, \tag{67}$$

so ist leicht zu sehen, dass sie genau eine Wurzel im Intervall (63) besitzt. Lässt man nämlich ω von 0 bis 1 wandern, so wächst die linke Seite von (67) monoton von 0 bis 1 und die rechte Seite fällt monoton von 1 bis 0. Hat man ω bestimmt, so erhält man ρ vermöge (64) und dann $\hat{p}_1$, $\hat{p}_2$ aus (62), während $\hat{p}_3=0$ ist. Aus (37) folgt jetzt

$$\tfrac{2}{9}\lambda=\{m_2m_3\omega^{-1}+m_3m_1+m_1m_2(1-\omega)^{-1}\}\rho^{-1}. \tag{68}$$

Man erhält die beiden anderen geradlinigen Fälle, wenn man in (64) und (66) die Massen m_1, m_2, m_3 zyklisch vertauscht.

Aus (59), (60), (62), (64), (66) ist nun ersichtlich, dass sowohl im gleichseitigen Fall als auch in den drei geradlinigen Fällen die Grössen $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$ eindeutig durch m_1, m_2, m_3 bestimmt sind. Wir hatten bisher eine solche Folge $t\to 0$ betrachtet, für welche die Werte p_k^* $(k=1,2,3)$ konvergieren. Da aber die p_k^* für $t>0$ stetige Funktionen von t sind und nur jene vier isolierten Systeme von Häufungswerten $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$ möglich sind, so konvergieren die p_k^* auch, wenn t beliebig gegen 0 strebt. Damit ist bewiesen, dass die Ausdrücke $r_kt^{-2/3}$ $(k=1,2,3)$ für $t\to 0$ gegen positive Grenzwerte streben, nämlich entweder gegen den durch (60) festgelegten Wert r des gleichseitigen Falles oder gegen die Werte $\omega\rho$, ρ, $(1-\omega)\rho$ der drei geradlinigen Fälle, welche sich aus (64), (66) bestimmen, nach etwaiger zyklischer Vertauschung der Indizes.

Wir wollen weiterhin von den 3 geradlinigen Fällen nur noch den durch (64), (66) fixierten studieren, da die beiden anderen durch Vertauschung der Indizes auf diesen zurückgeführt werden.

7. Ein Spezialfall

Bei den bekannten von Lagrange entdeckten speziellen Lösungen des Dreikörperproblemes bewegen sich A_1, A_2, A_3 auf drei in einer Ebene gelegenen Kegelschnitten, während das Dreieck $A_1A_2A_3$ dauernd einem festen Dreieck ähnlich bleibt. Dabei ergeben sich für die Form des Dreiecks zwei Möglichkeiten: Entweder bilden A_1, A_2, A_3 die Ecken eines gleichseitigen Dreiecks oder sie liegen auf einer Geraden. Setzt man im letzteren Fall voraus, dass A_2 zwischen A_1 und A_3 gelegen ist, so erhält man als Wert des Verhältnisses der Strecken A_2A_3 und A_1A_3 gerade die durch (66) definierte Zahl ω. Hierdurch wird nahe gelegt, die unter den Lagrangeschen Lösungen enthaltenen Dreierstossbahnen aufzusuchen, um dann die allgemeinen Dreierstossbahnen mit diesen vergleichen zu können.

Man erhält die Kollisionsbahnen unter den Lagrangeschen Lösungen, indem man die kegelschnittförmige Bewegung der Massenpunkte in eine geradlinige ausarten lässt. Dementsprechend machen wir den speziellen Ansatz

$$x_k = \hat{x}_k g(t), \qquad y_k = \hat{y}_k g(t) \qquad (k = 1, 2, 3)$$

mit konstanten $\hat{x}_k$, $\hat{y}_k$ und einer zweimal differentiierbaren Funktion $g(t)$, die für $t > 0$ positiv ist und für $t = 0$ verschwindet. Versteht man unter $\hat{U}$ den Wert von U mit $\hat{x}_k$, $\hat{y}_k$ anstelle von x_k, y_k, so gehen die Differentialgleichungen (2) über in

$$(69) \qquad m_k \hat{w}_k \ddot{g} g^2 = \frac{\partial \hat{U}}{\partial \hat{w}_k} \qquad (\hat{w}_k = \hat{x}_k, \hat{y}_k;\ k = 1, 2, 3).$$

Folglich ist der Ausdruck $\ddot{g}g^2$ konstant, aber nicht 0, weil

$$\sum_{k=1}^{3} \left(\hat{x}_k \frac{\partial \hat{U}}{\partial \hat{x}_k} + \hat{y}_k \frac{\partial \hat{U}}{\partial \hat{y}_k} \right) = -\hat{U} \neq 0$$

ist. Dann kann man aber die Normierung

$$(70) \qquad \ddot{g}g^2 = -\tfrac{2}{9}$$

treffen und auf die Gleichungen (69) die Überlegungen anwenden, mit denen wir im vorigen Paragraphen die Relationen (54) untersucht haben. Wegen der orthogonalen Invarianz kann man noch $\hat{x}_1 > \hat{x}_3$, $\hat{y}_1 = \hat{y}_3$, $\hat{y}_2 \geqq 0$ annehmen und erhält für die Punkte $(\hat{x}_k, \hat{y}_k)$ $(k = 1, 2, 3)$ genau die Punkte $\hat{A}_1$, $\hat{A}_2$, $\hat{A}_3$ des gleichseitigen oder des geradlinigen Falles. Für die relativen Koordinaten $\hat{x}_1 - \hat{x}_3 = \hat{p}_1$, $\hat{x}_2 - \hat{x}_3 = \hat{p}_2$, $\hat{y}_2 - \hat{y}_3 = \hat{p}_3$ gilt dann nach (59) im gleichseitigen Falle

$$\hat{p}_1 = r, \qquad \hat{p}_2 = \tfrac{1}{2}r, \qquad \hat{p}_3 = \tfrac{1}{2}r\sqrt{3}$$

und nach (62) im geradlinigen Falle

$$\hat{p}_1 = \rho, \qquad \hat{p}_2 = \omega\rho, \qquad \hat{p}_3 = 0.$$

wobei r, ω, ρ durch (60), (66), (64) festgelegt werden.

Die Integration von (70) ergibt

$$\dot{g}^2 = \tfrac{4}{9}g^{-1} + c$$

mit konstantem c. Wählt man speziell $c = 0$, so erhält man durch nochmalige Integration

$$g = t^{2/3}$$

und damit als spezielle Dreierstosslösung

$$x_k = \hat{x}_k t^{2/3}, \qquad y_k = \hat{y}_k t^{2/3} \qquad (k = 1, 2, 3).$$

Für einen späteren Zweck berechnen wir noch die zu dieser Lösung gehörigen

Werte von $\dot{x}_1$, $\dot{x}_2$, $\dot{y}_2$. Nach (6) ist

$$\hat{x}_1 = \left(1 - \frac{m_1}{m}\right)\hat{p}_1 - \frac{m_2}{m}\hat{p}_2, \qquad \hat{x}_2 = \left(1 - \frac{m_2}{m}\right)\hat{p}_2 - \frac{m_1}{m}\hat{p}_1,$$

$$\hat{y}_2 = \left(1 - \frac{m_2}{m}\right)\hat{p}_3,$$

also im gleichseitigen Falle

$$(71) \qquad \dot{x}_1 = \frac{m_2 + 2m_3}{3m} rt^{-1/3}, \qquad \dot{x}_2 = \frac{m_3 - m_1}{3m} rt^{-1/3}, \qquad \dot{y}_2 = \frac{m_1 + m_3}{m\sqrt{3}} rt^{-1/3}$$

und im geradlinigen Falle

$$(72) \quad \dot{x}_1 = 2\frac{m_2(1-\omega) + m_3}{3m}\rho t^{-1/3}, \quad \dot{x}_2 = 2\frac{m_3\omega - m_1(1-\omega)}{3m}\rho t^{-1/3}, \quad \dot{y}_2 = 0.$$

8. Reduktion der Differentialgleichungen

Um das Verhalten der Dreierstosslösungen bei $t = 0$ noch näher zu untersuchen, müssen wir die in (55) definierten Grössen p_1, p_2, p_3 in die Differentialgleichungen (2) einführen. Zunächst bilden wir die Relativkoordinaten von A_1 und A_2 in bezug auf A_3 im ruhenden Koordinatensystem

$$(73) \quad \xi_1 = x_1 - x_3, \qquad \xi_2 = y_1 - y_3, \qquad \xi_3 = x_2 - x_3, \qquad \xi_4 = y_2 - y_3$$

und setzen noch

$$(74) \qquad m_1\dot{x}_1 = \eta_1, \qquad m_1\dot{y}_1 = \eta_2, \qquad m_2\dot{x}_2 = \eta_3, \qquad m_2\dot{y}_2 = \eta_4.$$

Nach (6) ist dann

$$(75) \qquad m_3\dot{x}_3 = -(\eta_1 + \eta_3), \qquad m_3\dot{y}_3 = -(\eta_2 + \eta_4),$$

also nach (7)

$$(76) \qquad T = \frac{1}{2m_1}(\eta_1^2 + \eta_2^2) + \frac{1}{2m_2}(\eta_3^2 + \eta_4^2) + \frac{1}{2m_3}\{(\eta_1 + \eta_3)^2 + (\eta_2 + \eta_4)^2\}.$$

Ferner wird

$$(77) \quad r_1^2 = \xi_3^2 + \xi_4^2, \qquad r_2^2 = \xi_1^2 + \xi_2^2, \qquad r_3^2 = (\xi_1 - \xi_3)^2 + (\xi_2 - \xi_4)^2,$$

also U eine Funktion von ξ_1, ξ_2, ξ_3, ξ_4 allein. Die Differentialgleichungen (2) gehen dann über in das System achter Ordnung

$$(78) \quad \begin{cases} \dot{\xi}_1 = \left(\frac{1}{m_1} + \frac{1}{m_3}\right)\eta_1 + \frac{1}{m_3}\eta_3, & \dot{\xi}_3 = \left(\frac{1}{m_2} + \frac{1}{m_3}\right)\eta_3 + \frac{1}{m_3}\eta_1, \\ \dot{\xi}_2 = \left(\frac{1}{m_1} + \frac{1}{m_3}\right)\eta_2 + \frac{1}{m_3}\eta_4, & \dot{\xi}_4 = \left(\frac{1}{m_2} + \frac{1}{m_3}\right)\eta_4 + \frac{1}{m_3}\eta_2, \\ \qquad \dot{\eta}_k = \frac{\partial U}{\partial \xi_k} & (k = 1, \cdots, 4). \end{cases}$$

Führt man die Energie

$$E = T - U$$

ein, so hat das System (78) die kanonische Form

$$\dot{\xi}_k = \frac{\partial E}{\partial \eta_k}, \qquad \dot{\eta}_k = -\frac{\partial E}{\partial \xi_k} \qquad (k = 1, \cdots, 4). \tag{79}$$

Um nun die Differentialgleichungen für die Relativkoordinaten im bewegten Koordinatensystem aus §6 aufzustellen, macht man am bequemsten von der Jacobischen Transformationstheorie Gebrauch. Nach dieser gilt bekanntlich folgender Satz:

Es sei H eine Funktion von $2n$ Variabeln η_k, p_k $(k = 1, \cdots, n)$ mit stetigen partiellen Ableitungen zweiter Ordnung in der Umgebung einer Stelle, an welcher die n-reihige Determinante

$$D = \left| \frac{\partial^2 H}{\partial p_k \, \partial \eta_l} \right| \neq 0 \tag{80}$$

ist. Dann wird durch den Ansatz

$$\xi_k = \frac{\partial H}{\partial \eta_k}, \qquad q_k = \frac{\partial H}{\partial p_k} \qquad (k = 1, \cdots, n) \tag{81}$$

eine Variabelntransformation definiert, welche das Hamiltonsche System

$$\dot{\xi}_k = \frac{\partial E}{\partial \eta_k}, \qquad \dot{\eta}_k = -\frac{\partial E}{\partial \xi_k} \qquad (k = 1, \cdots, n)$$

in das Hamiltonsche System

$$\dot{p}_k = \frac{\partial E}{\partial q_k}, \qquad \dot{q}_k = -\frac{\partial E}{\partial p_k} \qquad (k = 1, \cdots, n)$$

überführt.

Bedeutet p_4 den Winkel zwischen der ruhenden Abszissenachse und der Richtung A_3A_1, so bestehen zwischen ξ_1, ξ_2, ξ_3, ξ_4 und den in (55) erklärten Relativkoordinaten p_1, 0, p_2, p_3 im bewegten Koordinatensystem die Gleichungen

$$\begin{gathered} \xi_1 = p_1 \cos p_4, \qquad \xi_2 = p_1 \sin p_4, \qquad \xi_3 = p_2 \cos p_4 - p_3 \sin p_4, \\ \xi_4 = p_2 \sin p_4 + p_3 \cos p_4. \end{gathered} \tag{82}$$

Wir wählen

$$\begin{aligned} H = \eta_1 p_1 \cos p_4 + \eta_2 p_1 \sin p_4 &+ \eta_3 (p_2 \cos p_4 - p_3 \sin p_4) \\ &+ \eta_4 (p_2 \sin p_4 + p_3 \cos p_4) \end{aligned}$$

und wenden den Transformationssatz für $n = 4$ an. Eine leichte Rechnung ergibt für die Determinante D in (80) den Wert p_1, und dieser ist von 0 ver-

schieden, solange keine Kollision vorliegt, also für $t > 0$. Die erste Gleichung in (81) ist wegen (82) erfüllt, die zweite Gleichung ergibt die Formeln

$$(83)\qquad \begin{gathered} q_1 = \eta_1 \cos p_4 + \eta_2 \sin p_4, \qquad q_2 = \eta_3 \cos p_4 + \eta_4 \sin p_4, \\ q_3 = -\eta_3 \sin p_4 + \eta_4 \cos p_4, \end{gathered}$$

$$(84)\qquad \begin{aligned} q_4 = -\eta_1 p_1 \sin p_4 + \eta_2 p_1 \cos p_4 - \eta_3(p_2 \sin p_4 + p_3 \cos p_4) \\ + \eta_4(p_2 \cos p_4 - p_3 \sin p_4), \end{aligned}$$

also

$$(85)\qquad q_4 = p_1(-\eta_1 \sin p_4 + \eta_2 \cos p_4) + p_2 q_3 - p_3 q_2 .$$

Setzt man zur Abkürzung noch

$$(86)\qquad q_0 = (p_3 q_2 - p_2 q_3 + q_4) p_1^{-1},$$

so ist nach (85)

$$q_0 = -\eta_1 \sin p_4 + \eta_2 \cos p_4 ,$$

und in Verbindung mit (83) folgt

$$(87)\qquad \begin{cases} \eta_1 = q_1 \cos p_4 - q_0 \sin p_4, & \eta_2 = q_1 \sin p_4 + q_0 \cos p_4, \\ \eta_3 = q_2 \cos p_4 - q_3 \sin p_4, & \eta_4 = q_2 \sin p_4 + q_3 \cos p_4. \end{cases}$$

Daher gilt nach (76)

$$(88)\qquad T = \frac{1}{2m_1}(q_0^2 + q_1^2) + \frac{1}{2m_2}(q_2^2 + q_3^2) + \frac{1}{2m_3}\{(q_0 + q_3)^2 + (q_1 + q_2)^2\},$$

ferner nach (77) und (82)

$$(89)\qquad r_1^2 = p_2^2 + p_3^2, \qquad r_2 = p_1, \qquad r_3^2 = (p_1 - p_2)^2 + p_3^2 .$$

Auf Grund des Transformationssatzes geht das System (79) durch die Substitutionen (82), (87) über in

$$(90)\qquad \dot{p}_k = \frac{\partial E}{\partial q_k}, \qquad \dot{q}_k = -\frac{\partial E}{\partial p_k} \qquad (k = 1, \cdots, 4),$$

wobei $E = T - U$ nach (88) und (89) als Funktion der p_k, q_k $(k = 1, \cdots, 4)$ anzusehen ist. Mit Rücksicht auf (86) ist ersichtlich, dass E die Variable p_4 nicht enthält, also

$$\frac{\partial E}{\partial p_4} = 0$$

ist. Die zweite Gleichung (90) ergibt für $k = 4$, dass q_4 konstant ist. Dies ist gerade die Aussage des Flächenintegrals, denn aus (73), (74), (75), (82), (84) folgt

$$q_4 = -\eta_1 \xi_2 + \eta_2 \xi_1 - \eta_3 \xi_4 + \eta_4 \xi_3 = \sum_{k=1}^{3} m_k(x_k \dot{y}_k - y_k \dot{x}_k);$$

also ist q_4 die Konstante γ_3 aus (5). Nach §5 ist daher für jede Dreierstossbahn

$$q_4 = 0, \tag{91}$$

und wir erhalten für p_k, q_k ($k = 1, 2, 3$) das System sechster Ordnung

$$\dot{p}_k = \left(\frac{\partial E}{\partial q_k}\right)_{q_4=0}, \qquad \dot{q}_k = -\left(\frac{\partial E}{\partial p_k}\right)_{q_4=0} \qquad (k = 1, 2, 3).$$

Ist dieses integriert, so ergibt sich p_4 durch Quadratur aus

$$\dot{p}_4 = \left(\frac{\partial E}{\partial q_4}\right)_{q_4=0}.$$

9. Das asymptotische Verhalten der p_k und q_k

Wir haben in §6 bewiesen, dass die Ausdrücke $p_k^* = p_k t^{-2/3}$ ($k = 1, 2, 3$) für $t \to 0$ Grenzwerte $\hat{p}_k$ haben. Nach (59) ist im gleichseitigen Falle

$$p_1 \sim r t^{2/3}, \qquad p_2 \sim \tfrac{1}{2} r t^{2/3}, \qquad p_3 \sim \frac{\sqrt{3}}{2} r t^{2/3}, \tag{92}$$

wobei r durch (60) gegeben ist, und nach (62) im geradlinigen Falle

$$p_1 \sim \rho t^{2/3}, \qquad p_2 \sim \omega\rho t^{2/3}, \qquad p_3 = o(t^{2/3}), \tag{93}$$

wobei die Zahlen ω, ρ in (66), (64) festgelegt sind.

Wir wenden uns jetzt zur Untersuchung des asymptotischen Verhaltens von q_1, q_2, q_3. Schreibt man zur Abkürzung

$$\cos p_4 = \mu, \qquad \sin p_4 = \nu, \tag{94}$$

so ist nach (74), (75), (87)

$$\begin{cases} m_1\dot{x}_1 = q_1\mu - q_0\nu, \qquad m_2\dot{x}_2 = q_2\mu - q_3\nu, \\ \qquad m_3\dot{x}_3 = -(q_1 + q_2)\mu + (q_0 + q_3)\nu, \\ m_1\dot{y}_1 = q_1\nu + q_0\mu, \qquad m_2\dot{y}_2 = q_2\nu + q_3\mu, \\ \qquad m_3\dot{y}_3 = -(q_1 + q_2)\nu - (q_0 + q_3)\mu. \end{cases} \tag{95}$$

Nach (6), (73), (82) ist ferner

$$\begin{cases} mx_1 = \{(m_2 + m_3)p_1 - m_2p_2\}\mu + m_2p_3\nu, \\ \qquad\qquad my_1 = \{(m_2 + m_3)p_1 - m_2p_2\}\nu - m_2p_3\mu, \\ mx_2 = \{(m_1 + m_3)p_2 - m_1p_1\}\mu - (m_1 + m_3)p_3\nu, \\ \qquad\qquad my_2 = \{(m_1 + m_3)p_2 - m_1p_1\}\nu + (m_1 + m_3)p_3\mu, \\ mx_3 = -(m_1p_1 + m_2p_2)\mu + m_2p_3\nu, \quad my_3 = -(m_1p_1 + m_2p_2)\nu - m_2p_3\mu. \end{cases} \tag{96}$$

Hieraus folgt mit Rücksicht auf (10)

$$\tfrac{1}{2}\dot{J} = p_1q_1 + p_2q_2 + p_3q_3\,,$$

also nach (25)

$$p_1q_1 + p_2q_2 + p_3q_3 \sim \tfrac{2}{3}\lambda t^{1/3}, \tag{97}$$

wobei λ im gleichseitigen Falle durch (61), im geradlinigen Falle durch (68) festgelegt wird. Ausserdem ist nach (38)

$$x_1\dot{y}_1 - y_1\dot{x}_1 = o(t^{1/3}), \qquad x_2\dot{y}_2 - y_2\dot{x}_2 = o(t^{1/3}),$$

also

$$\{(m_2 + m_3)p_1 - m_2p_2\}q_0 + m_2p_3q_1 = o(t^{1/3}), \tag{98}$$

$$\{(m_1 + m_3)p_2 - m_1p_1\}q_3 - (m_1 + m_3)p_3q_2 = o(t^{1/3}). \tag{99}$$

Zufolge (43), (83), (86) ist

$$q_k = O(t^{-1/3}) \qquad (k = 0, \cdots, 3). \tag{100}$$

Im gleichseitigen Falle folgen aus (91), (92), (97), (98), (99), (100) die Beziehungen

$$2q_1 + q_2 + q_3\sqrt{3} \sim \tfrac{4}{3}\lambda r^{-1}t^{-1/3}, \tag{101}$$

$$(m_2 + 2m_3)(q_2\sqrt{3} - q_3) + 2m_2q_1\sqrt{3} = o(t^{-1/3}), \tag{102}$$

$$(m_3 - m_1)q_3 - (m_1 + m_3)q_2\sqrt{3} = o(t^{-1/3}). \tag*{(102)}$$

Setzt man

$$q_3 = \frac{1}{\sqrt{3}}\, m_2(m_1 + m_3)q,$$

so wird nach (103)

$$q_2 = \tfrac{1}{3}m_2(m_3 - m_1)q + o(t^{-1/3})$$

und nach (102)

$$q_1 = \tfrac{1}{3}m_1(m_2 + 2m_3)q + o(t^{-1/3}),$$

also nach (101)

$$\tfrac{4}{3}(m_2m_3 + m_3m_1 + m_1m_2)q \sim \tfrac{4}{3}\lambda r^{-1}t^{-1/3},$$

woraus nach (60), (61)

$$q \sim \frac{r}{m}\, t^{-1/3}$$

und folglich

$$\text{(104)}\qquad \begin{gathered} q_1 \sim \frac{m_1}{3m}(m_2 + 2m_3)rt^{-1/3}, \qquad q_2 \sim \frac{m_2}{3m}(m_3 - m_1)rt^{-1/3}, \\ q_3 \sim \frac{m_2}{m\sqrt{3}}(m_1 + m_3)rt^{-1/3} \end{gathered}$$

sich ergibt.

Im geradlinigen Falle ist nach (86), (91), (93), (98), (100)

$$-\{m_2(1-\omega) + m_3\}\omega q_3 = o(t^{-1/3}),$$

also

$$\text{(105)}\qquad q_3 = o(t^{-1/3}), \qquad q_0 = o(t^{-1/3}).$$

Nach (38) ist ferner

$$m_1 m_2(x_1\dot{x}_2 - x_2\dot{x}_1 + y_1\dot{y}_2 - y_2\dot{y}_1) = o(t^{1/3}),$$

also nach (95), (96), (105)

$$\text{(106)}\qquad m_1\{m_2(1-\omega) + m_3\}q_2 + m_2\{m_1(1-\omega) - m_3\omega\}q_1 = o(t^{1/3}).$$

Setzt man diesmal

$$q_1 = m_1\{m_2(1-\omega) + m_3\}q,$$

so wird

$$q_2 = m_2\{m_3\omega - m_1(1-\omega)\}q + o(t^{1/3})$$

und nach (97)

$$\text{(107)}\qquad \{m_2m_3\omega^2 + m_3m_1 + m_1m_2(1-\omega)^2\}q \sim \tfrac{2}{3}\lambda\rho^{-1}t^{-1/3}.$$

Multipliziert man (64) mit $m_1\{m_2(1-\omega) + m_3\}$ und (65) mit $m_2\{m_3\omega - m_1(1-\omega)\}$, so folgt durch Addition

$$\tfrac{2}{9}\{m_2m_3\omega^2 + m_3m_1 + m_1m_2(1-\omega)^2\}\rho^3 = \{m_2m_3\omega^{-1} + m_3m_1 + m_1m_2(1-\omega)^{-1}\}m.$$

Benutzt man noch (68), so geht (107) über in

$$q \sim \frac{2}{3}\frac{\rho}{m}t^{-1/3}.$$

Demnach ist jetzt

$$\text{(108)}\qquad \begin{gathered} q_1 \sim \frac{2m_1}{3m}\{m_2(1-\omega) + m_3\}\rho t^{-1/2}, \\ q_2 \sim \frac{2m_2}{3m}\{m_3\omega - m_1(1-\omega)\}\rho t^{-1/3}, \qquad q_3 = o(t^{-1/3}). \end{gathered}$$

Durch (104) und (108) ist das asymptotische Verhalten von q_1, q_2, q_3 im gleichseitigen und im geradlinigen Falle festgestellt. Dass auch p_4 einen Grenzwert für $t \to 0$ hat, wird sich erst im späteren Verlauf der Untersuchung ergeben.

Man kann (104) auch mit etwas geringerer Rechnung aus den Formeln (101), (102), (103) erhalten, indem man von §7 Gebrauch macht. Da nämlich die auf den linken Seiten dieser Formeln stehenden linearen Formen von q_1, q_2, q_3 linear unabhängig sind, so sind die Grenzwerte von $q_k t^{1/3}$ $(k = 1, 2, 3)$ für $t \to 0$ eindeutig bestimmt. Also kann man sie durch Betrachtung des speziellen Falles von §7 ermitteln. Dort ist aber $p_4 = 0$ und folglich nach (95)

$$q_1 = m_1\dot{x}_1, \qquad q_2 = m_2\dot{x}_2, \qquad q_3 = m_2\dot{y}_2.$$

Aus (71) folgt dann (104). Analog kann man im geradlinigen Falle aus (72), (97), (105), (106) ohne weitere Rechnung auf (108) schliessen.

10. Die charakteristische Gleichung

Wir machen in den Differentialgleichungen (90) die Substitutionen

$$\text{(109)} \qquad \begin{cases} p_k = \overset{*}{p_k}\, t^{2/3}, & q_k = \overset{*}{q_k}\, t^{-1/3} \qquad (k = 1, 2, 3), \\ p_4 = \overset{*}{p_4}, & q_4 = \overset{*}{q_4}\, t^{1/3}, \end{cases}$$

$$\text{(110)} \qquad t = e^{-s}.$$

Nach (86), (88), (89) ist E eine Funktion von p_k, q_k $(k = 1, \cdots, 4)$, in welcher p_4 nicht auftritt. Ersetzt man darin die Variabeln p_k, q_k durch $\overset{*}{p_k}$, $\overset{*}{q_k}$, so möge E^* entstehen, und zwar ist

$$E^* = Et^{2/3}.$$

Das System (90) geht dadurch über in

$$\text{(111)} \qquad \begin{cases} \dfrac{d\overset{*}{p_k}}{ds} = \tfrac{2}{3}\overset{*}{p_k} - \dfrac{\partial E^*}{\partial \overset{*}{q_k}}, & \dfrac{d\overset{*}{q_k}}{ds} = -\tfrac{1}{3}\overset{*}{q_k} + \dfrac{\partial E^*}{\partial \overset{*}{p_k}} \qquad (k = 1, 2, 3), \\ \dfrac{d\overset{*}{p_4}}{ds} = -\dfrac{\partial E^*}{\partial \overset{*}{q_4}}, & \dfrac{d\overset{*}{q_4}}{ds} = \tfrac{1}{3}\overset{*}{q_4} + \dfrac{\partial E}{\partial \overset{*}{p_4}}. \end{cases}$$

Für $t \to 0$ ist $s \to \infty$. Wie im vorigen Paragraphen gezeigt wurde, haben bei diesem Grenzübergang die Ausdrücke $\overset{*}{p_k}$ und $\overset{*}{q_k}$ $(k = 1, 2, 3)$ bestimmte Grenzwerte $\hat{p}_k$ und $\hat{q}_k$. Im gleichseitigen Falle ist nach (92) und (104)

$$\text{(112)} \qquad \begin{cases} \hat{p}_1 = r, \qquad \hat{p}_2 = \tfrac{1}{2}r, \qquad \hat{p}_3 = \tfrac{1}{2}r\sqrt{3}, \\ \hat{q}_1 = \dfrac{m_1}{3m}(m_2 + 2m_3)r, \qquad \hat{q}_2 = \dfrac{m_2}{3m}(m_3 - m_1), \\ \hat{q}_3 = \dfrac{m_2}{m\sqrt{3}}(m_1 + m_3)r, \end{cases}$$

im geradlinigen Falle nach (93) und (108)

$$(113)\begin{cases} \hat{p}_1 = \rho, \qquad \hat{p}_2 = \omega\rho, \qquad \hat{p}_3 = 0, \\ \hat{q}_1 = \dfrac{2m_1}{3m}\{m_2(1-\omega) + m_3\}\rho, \qquad \hat{q}_2 = \dfrac{2m_2}{3m}\{m_3\omega - m_1(1-\omega)\}\rho, \\ \hat{q}_3 = 0. \end{cases}$$

Nach §7 kennen wir eine spezielle Dreierstosslösung, nämlich

$$p_k = \hat{p}_k t^{2/3}, \qquad q_k = \hat{q}_k t^{-1/3} \; (k = 1, 2, 3), \qquad p_4 = 0, \qquad q_4 = 0,$$

wobei also $\hat{p}_k$, $\hat{q}_k$ $(k = 1, 2, 3)$ im gleichseitigen Falle durch (112), im geradlinigen Falle durch (113) gegeben sind. Wir erhalten also eine spezielle Lösung des Systems (111), eine Gleichgewichtslösung, wenn wir p_k^*, q_k^* konstant gleich $\hat{p}_k, \hat{q}_k$ $(k = 1, 2, 3)$ und $p_4 = 0$, $q_4 = 0$ setzen. Um sämtliche Dreierstossbahnen zu bekommen, haben wir sämtliche Lösungen von (111) zu untersuchen, für welche p_k^*, q_k^* $(k = 1, 2, 3)$ die Grenzwerte $\hat{p}_k$, $\hat{q}_k$ haben, wenn s über alle Grenzen wächst.

Wir setzen noch

$$(114) \qquad p_k^* = \hat{p}_k + \delta_k, \quad q_k^* = \hat{q}_k + \delta_{k+3} \; (k = 1, 2, 3), \quad p_4^* = \delta_8, \quad q_4^* = \delta_7.$$

Für genügend kleine Werte der absoluten Beträge von δ_1, δ_2, δ_3 lässt sich E^* in eine Reihe nach Potenzen von $\delta_1, \cdots, \delta_7$ entwickeln, deren Koeffizienten noch von m_1, m_2, m_3 abhängen. Die Differentialgleichungen (111) gehen dadurch über in ein System der Gestalt

$$(115) \qquad \frac{d\delta_k}{ds} = \sum_{l=1}^{8} a_{kl}\delta_l + \varphi_k \qquad (k = 1, \cdots, 8),$$

wo a_{kl} $(k, l = 1, \cdots, 8)$ Konstante bedeuten und $\varphi_1, \cdots, \varphi_8$ Potenzreihen in $\delta_1, \cdots, \delta_8$ ohne konstante und lineare Glieder. Da E^* nicht die Variable p_4^* enthält, so gilt nach (111)

$$(116) \qquad a_{k8} = 0 \; (k = 1, \cdots, 8), \quad a_{77} = \tfrac{1}{3}, \quad a_{7l} = 0 \; (l \neq 7), \quad \varphi_7 = 0.$$

Es bedeute $\mathfrak{M}$ die achtreihige Matrix aus den Elementen a_{kl}, ferner $\mathfrak{A}$ die Untermatrix von $\mathfrak{M}$, die durch Streichung der beiden letzten Zeilen und Spalten aus $\mathfrak{M}$ hervorgeht. Wird eine Einheitsmatrix mit $\mathfrak{E}$ bezeichnet, so sind zufolge (116) die charakteristischen Polynome

$$G(z) = |z\mathfrak{E} - \mathfrak{M}|, \qquad F(z) = |z\mathfrak{E} - \mathfrak{A}|$$

der Matrizen $\mathfrak{M}$ und $\mathfrak{A}$ durch die Gleichung

$$(117) \qquad G(z) = z(z - \tfrac{1}{3})F(z)$$

verknüpft.

Für jede Dreierstosslösung ist $q_4 = 0$, also $\delta_7 = 0$. Es sei ψ_k der Wert von φ_k für $\delta_7 = 0$. Wir haben dann das System

$$\frac{d\delta_k}{ds} = \sum_{l=1}^{6} a_{kl}\delta_l + \psi_k \qquad (k = 1, \cdots, 6), \tag{118}$$

$$\frac{d\delta_8}{ds} = \sum_{l=1}^{6} a_{8l}\delta_l + \psi_8 \tag{119}$$

zu lösen, unter der Bedingung $\delta_k \to 0$ $(k = 1, \cdots, 6)$ für $s \to \infty$. Auf den rechten Seiten dieser 7 Differentialgleichungen tritt δ_8 nicht auf. Hat man also das System (118) vollständig gelöst, unter jener Bedingung, so ergibt sich $\delta_8 = p_4$ aus (119) durch eine Quadratur. Zur näheren Diskussion von (118) ist die Kenntnis der charakteristischen Wurzeln der Matrix $\mathfrak{A}$ notwendig. Die direkte Berechnung der Determinante $| z\mathfrak{E} - \mathfrak{A} |$ ist recht mühsam, da die Koeffizienten a_{kl} sich nicht bequem bestimmen lassen. Einfacher erhält man $F(z)$ durch folgende Überlegung:

Bezeichnet man die rechten Seiten der Differentialgleichungen (115) mit Φ_k $(k = 1, \cdots, 8)$, so hat man das System

$$\frac{d\delta_k}{ds} = \Phi_k \qquad (k = 1, \cdots, 8) \tag{120}$$

und $\mathfrak{M}$ ist die Funktionalmatrix $\left(\frac{\partial\Phi_k}{\partial\delta_l}\right)$ an der Stelle $\delta_1 = 0, \cdots, \delta_8 = 0$. Man betrachte jetzt $\delta_1, \cdots, \delta_8$ als zweimal stetig differentiierbare Funktionen von 8 neuen Variabeln $\theta_1, \cdots, \theta_8$, und zwar möge für das Wertsystem $\theta_k = \hat{\theta}_k$ $(k = 1, \cdots, 8)$ speziell $\delta_k = 0$ $(k = 1, \cdots, 8)$ sein und die Funktionaldeterminante der δ_k bezüglich der Variabeln θ_k an der Stelle $\theta_k = \hat{\theta}_k$ nicht verschwinden. Durch diese Transformation gehen die Differentialgleichungen (120) über in

$$\frac{d\theta_k}{ds} = \sum_{g=1}^{8} \frac{\partial\theta_k}{\partial\delta_g}\Phi_g \qquad (k = 1, \cdots, 8),$$

wobei die rechten Seiten als Funktionen der θ_k anzusehen sind. Bezeichnet man diese rechten Seiten zur Abkürzung mit Ψ_k $(k = 1, \cdots, 8)$, so ist

$$\frac{\partial\Psi_k}{\partial\theta_l} = \sum_{g=1}^{8} \Phi_g \frac{\partial}{\partial\theta_l}\left(\frac{\partial\theta_k}{\partial\delta_g}\right) + \sum_{g,h=1}^{8} \frac{\partial\theta_k}{\partial\delta_g}\frac{\partial\Phi_g}{\partial\delta_h}\frac{\partial\delta_h}{\partial\theta_l}. \tag{121}$$

Es bedeute $\mathfrak{P}$ die Funktionalmatrix $\left(\frac{\partial\Psi_k}{\partial\theta_l}\right)$ an der Stelle $\theta_1 = \hat{\theta}_1, \cdots, \theta_8 = \hat{\theta}_8$ und $\mathfrak{G}$ die Funktionalmatrix $\left(\frac{\partial\delta_k}{\partial\theta_l}\right)$ an derselben Stelle. Da der erste Summand auf der rechten Seite von (121) an dieser Stelle verschwindet, so ergibt (121) die Beziehung

$$\mathfrak{P} = \mathfrak{G}^{-1}\mathfrak{M}\mathfrak{G}$$

und folglich ist

$$| z\mathfrak{E} - \mathfrak{P} | = | z\mathfrak{E} - \mathfrak{G}^{-1}\mathfrak{M}\mathfrak{G} | = | \mathfrak{G}^{-1}(z\mathfrak{E} - \mathfrak{M})\mathfrak{G} | = | z\mathfrak{E} - \mathfrak{M} | = G(z). \tag{122}$$

Um diese Formel zur Berechnung von $G(z)$ anzuwenden, setzen wir noch

$$(123)\qquad \xi_k = \xi_k^* t^{-2/3}, \qquad \eta_k = \eta_k^* t^{-1/3} \qquad (k = 1, \cdots, 4),$$

$$q_0 = q_0^* t^{-1/3}, \qquad \hat{q}_0 = (\hat{p}_3\hat{q}_2 - \hat{p}_2\hat{q}_3)\hat{p}_1^{-1}.$$

Nach (82), (86), (87), (109) ist dann

$$\xi_1^* = \mu p_1^*, \qquad \xi_2^* = \nu p_1^*, \qquad \xi_3^* = \mu p_2^* - \nu p_3^*, \qquad \xi_4^* = \nu p_2^* + \mu p_3^*,$$

$$\eta_1^* = \mu q_1^* - \nu q_0^*, \quad \eta_2^* = \nu q_1^* + \mu q_0^*, \quad \eta_3^* = \mu q_2^* - \nu q_3^*, \quad \eta_4^* = \nu q_2^* + \mu q_3^*,$$

wobei μ, ν in (94) erklärt sind. Zufolge (114) sind die ξ_k^*, η_k^* Funktionen von $\delta_1, \cdots, \delta_8$, die an der Stelle $\delta_1 = 0, \cdots, \delta_8 = 0$ die Werte

$$\hat{\xi}_1 = \hat{p}_1, \quad \hat{\xi}_2 = 0, \quad \hat{\xi}_3 = \hat{p}_2, \quad \hat{\xi}_4 = \hat{p}_3, \quad \hat{\eta}_1 = \hat{q}_1, \quad \hat{\eta}_2 = \hat{q}_0, \quad \hat{\eta}_3 = \hat{q}_2, \quad \hat{\eta}_4 = \hat{q}_3$$

haben. Endlich sei

$$(124)\begin{cases} \theta_k = \xi_k^* & (k = 1, \cdots, 4), \\ \theta_5 = -\left(\dfrac{1}{m_1} + \dfrac{1}{m_3}\right)\eta_1^* - \dfrac{1}{m_3}\eta_3^*, & \theta_6 = -\left(\dfrac{1}{m_1} + \dfrac{1}{m_3}\right)\eta_2^* - \dfrac{1}{m_3}\eta_4^*, \\ \theta_7 = -\left(\dfrac{1}{m_2} + \dfrac{1}{m_3}\right)\eta_3^* - \dfrac{1}{m_3}\eta_1^*, & \theta_8 = -\left(\dfrac{1}{m_2} + \dfrac{1}{m_3}\right)\eta_4^* - \dfrac{1}{m_3}\eta_2^*. \end{cases}$$

Durch eine einfache Rechnung erhält man für die Funktionaldeterminante der θ_k als Funktionen der δ_k den Wert $m^2(m_1m_2m_3)^{-2} \neq 0$. Vermöge der Substitutionen (123), (124) gehen nun die Differentialgleichungen (78) über in

$$(125)\qquad \frac{d\theta_k}{ds} = \tfrac{2}{3}\theta_k + \theta_{k+4}, \qquad \frac{d\theta_{k+4}}{ds} = -\tfrac{1}{3}\theta_{k+4} - G_k \qquad (k = 1, \cdots, 4)$$

mit

$$G_1 = m_2\theta_3R_1 + (m_1 + m_3)\theta_1R_2 + m_2(\theta_1 - \theta_3)R_3,$$

$$G_2 = m_2\theta_4R_1 + (m_1 + m_3)\theta_2R_2 + m_2(\theta_2 - \theta_4)R_3,$$

$$G_3 = m_1\theta_1R_1 + (m_2 + m_3)\theta_3R_2 + m_1(\theta_3 - \theta_1)R_3,$$

$$G_4 = m_1\theta_2R_1 + (m_2 + m_3)\theta_4R_2 + m_1(\theta_4 - \theta_2)R_3,$$

$$R_1 = (\theta_3^2 + \theta_4^2)^{-3/2}, \quad R_2 = (\theta_1^2 + \theta_2^2)^{-3/2}, \quad R_3 = \{(\theta_1 - \theta_3)^2 + (\theta_2 - \theta_4)^2\}^{-3/2}.$$

Für die Werte der Ableitungen

$$\frac{\partial G_k}{\partial \theta_l} = c_{kl} \qquad (k, l = 1, \cdots, 4)$$

an der Stelle

$$(126)\qquad \theta_1 = \hat{p}_1, \qquad \theta_2 = 0, \qquad \theta_3 = \hat{p}_2, \qquad \theta_4 = \hat{p}_3$$

findet man nach (112) im gleichseitigen Falle

$$r^3 c_{kl} = \begin{cases} \frac14 m_2 - 2(m_1 + m_3),\ \frac{3\sqrt{3}}{4}\, m_2,\ 0,\ -\frac{3\sqrt{3}}{2}\, m_2, \\ \frac{3\sqrt{3}}{4}\, m_2,\ m_1 + m_3 - \frac54 m_2,\ -\frac{3\sqrt{3}}{2}\, m_2,\ 0, \\ -\frac94 m_1,\ -\frac{3\sqrt{3}}{4}\, m_1,\ \frac14(m_1 + m_2 + m_3),\ \frac{3\sqrt{3}}{4}\,(m_1 - m_2 - m_3), \\ -\frac{3\sqrt{3}}{4}\, m_1,\ \frac94 m_1,\ \frac{3\sqrt{3}}{4}\,(m_1 - m_2 - m_3),\ -\frac54(m_1 + m_2 + m_3), \end{cases}$$

und nach (113) im geradlinigen Falle

$$\rho^3 c_{kl} = \begin{cases} -2(m_1 + m_3) - 2m_2\omega^{-3},\ 0,\ 2m_2\{\omega^{-3} - (1 - \omega)^{-3}\},\ 0, \\ 0,\ m_1 + m_3 + m_2\omega^{-3},\ 0,\ m_2\{(1 - \omega)^{-3} - \omega^{-3}\}, \\ -2m_1(1 - \omega^{-3}),\ 0,\ -2m_1\omega^{-3} - 2(m_2 + m_3)(1 - \omega)^{-3},\ 0, \\ 0,\ m_1(1 - \omega^{-3}),\ 0,\ m_1\omega^{-3} + (m_2 + m_3)(1 - \omega)^{-3}, \end{cases}$$

wo die Grössen r, ω, ρ durch (60), (66), (64) fixiert sind.

Bezeichnet man zur Abkürzung die vierreihige Matrix (c_{kl}) mit $\mathfrak{C}$, so hat die Funktionalmatrix der rechten Seiten von (125) als Funktionen von $\theta_1, \cdots, \theta_8$ an der Stelle (126) die Gestalt

$$\mathfrak{P} = \begin{pmatrix} \frac23\mathfrak{C} & \mathfrak{C} \\ -\mathfrak{C} & -\frac13\mathfrak{C} \end{pmatrix},$$

wo die rechts auftretenden Matrizen vierreihig sind, und nach (122) wird

$$G(z) = \begin{vmatrix} (z - \frac23)\mathfrak{C} & -\mathfrak{C} \\ \mathfrak{C} & (z + \frac13)\mathfrak{C} \end{vmatrix} = |\,(z + \tfrac13)(z - \tfrac23)\mathfrak{C} + \mathfrak{C}\,|.$$

Diese vierreihige Determinante bestimmt man durch direkte Rechnung unter Benutzung der angegebenen Werte der c_{kl}. Es ergibt sich ein einfaches Resultat. Setzt man

$$(z + \tfrac13)(z - \tfrac23) = x,$$

so wird im gleichseitigen Falle

$$G(z) = (x + \tfrac29)(x - \tfrac49)(x^2 - \tfrac29 x - \tfrac{8}{81} + \tfrac13 a) \tag{127}$$

mit

$$a = \frac{m_2 m_3 + m_3 m_1 + m_1 m_2}{(m_1 + m_2 + m_3)^2} \tag{128}$$

und im geradlinigen Falle

$$G(z) = (x + \tfrac{2}{9})(x - \tfrac{4}{9})(x + \tfrac{2}{9} + \tfrac{2}{9}b)(x - \tfrac{4}{9} - \tfrac{4}{9}b) \tag{129}$$

mit

$$b = \frac{m_1\{1 + (1-\omega)^{-1} + (1-\omega)^{-2}\} + m_3(1 + \omega^{-1} + \omega^{-2})}{m_1 + m_2\{\omega^{-2} + (1-\omega)^{-2}\} + m_3}. \tag{130}$$

Da

$$z(z - \tfrac{1}{3}) = x + \tfrac{2}{9}$$

ist, so erhalten wir nach (117) das charakteristische Polynom $F(z)$ der Matrix $\mathfrak{A}$, indem wir auf den rechten Seiten von (127) und (129) den Faktor $x + \frac{2}{9}$ fortlassen.

Die charakteristischen Wurzeln von $\mathfrak{A}$ sind also zufolge (127) im gleichseitigen Falle die 6 Zahlen

$$\left\{\begin{array}{ll} -a_0 = -\tfrac{2}{3}, & -a_1 = \tfrac{1}{6}\{1 - [13 + 12(1-3a)^{\frac{1}{2}}]^{\frac{1}{2}}\}, \\ -a_2 = \tfrac{1}{6}\{1 - [13 - 12(1-3a)^{\frac{1}{2}}]^{\frac{1}{2}}\}, & a_3 = \tfrac{1}{6}\{1 + [13 - 12(1-3a)^{\frac{1}{2}}]^{\frac{1}{2}}\}, \\ a_4 = \tfrac{1}{6}\{1 + [13 + 12(1-3a)^{\frac{1}{2}}]^{\frac{1}{2}}\}, & a_5 = 1. \end{array}\right. \tag{131}$$

Wegen der Beziehung

$$2m^2(1 - 3a) = (m_2 - m_3)^2 + (m_3 - m_1)^2 + (m_1 - m_2)^2$$

ist $a \leqq \frac{1}{3}$, und zwar $a = \frac{1}{3}$ nur für $m_1 = m_2 = m_3$; andererseits ist $a > 0$. Folglich sind die 6 Wurzeln sämtlich reell, und zwar $-a_0$, $-a_1$, $-a_2$ negativ, a_3, a_4, a_5 positiv. Ferner sind sie sämtlich verschieden, ausser im Falle $m_1 = m_2 = m_3$, wo $-a_1 = -a_2$ und $a_3 = a_4$ wird.

Im geradlinigen Falle ergeben sich aus (129) die Wurzeln

$$\left\{\begin{array}{lll} -b_0 = -\tfrac{2}{3}, & -b_1 = \tfrac{1}{6}[1 - (25 + 16b)^{\frac{1}{2}}], & b_2 = \tfrac{1}{6}[1 - (1 - 8b)^{\frac{1}{2}}], \\ b_3 = \tfrac{1}{6}[1 + (1 - 8b)^{\frac{1}{2}}], & b_4 = \tfrac{1}{6}[1 + (25 + 16b)^{\frac{1}{2}}], & b_5 = 1. \end{array}\right. \tag{132}$$

Von diesen sind $-b_0$ und $-b_1$ negativ, b_4 und b_5 positiv, b_2 und b_3 entweder positiv oder konjugiert komplex mit dem positiven Realteil $\frac{1}{6}$. Die Wurzeln sind alle verschieden, ausser im Falle $b = \frac{1}{8}$, wo $b_2 = b_3 = \frac{1}{6}$ wird. Die Gleichung $b = \frac{1}{8}$ liefert eine algebraische Bedingung für m_1, m_2, m_3, die nicht identisch erfüllt ist. Wählt man z.B. $m_1 = m_3$, so ergibt (66) den Wert $\omega = \frac{1}{2}$, und aus der Annahme $b = \frac{1}{8}$ folgt nach (130) die Bedingung $\dfrac{m_2}{m_1} = \dfrac{55}{4}$. Also sind im allgemeinen die Wurzeln sämtlich verschieden.

Dass die Werte $-\frac{2}{3}$ und 1 unter den Wurzeln auftreten, hätte man auch ohne Rechnung aus dem Energieintegral entnehmen können. Dagegen lässt sich die Bestimmung der übrigen Wurzeln wohl kaum einfacher durchführen, als es hier geschehen ist.

11. Asymptotische Bahnen

Wir betrachten ein System von Differentialgleichungen der Form

$$\frac{d\delta_k}{ds} = \sum_{l=1}^{n} a_{kl}\delta_l + \psi_k \qquad (k = 1, \cdots, n). \tag{133}$$

Hierin seien die a_{kl} reelle Konstante und die ψ_k Potenzreihen der Variabeln $\delta_1, \cdots, \delta_n$ mit reellen Koeffizienten, welche in einer gewissen Umgebung des Nullpunktes konvergieren und weder konstante noch lineare Glieder enthalten. Es seien $\lambda_1, \cdots, \lambda_n$ die charakteristischen Wurzeln der Matrix $(a_{kl}) = \mathfrak{A}$ und ρ_k der reelle Teil von λ_k $(k = 1, \cdots, n)$. Von diesen reellen Teilen seien p negativ und $n - p$ positiv, also keiner gleich 0. Wir denken uns die Wurzeln so angeordnet, dass

$$0 > \rho_1 \geqq \rho_2 \geqq \cdots \geqq \rho_p, \qquad \rho_{p+1} \geqq \rho_{p+2} \geqq \cdots \geqq \rho_n > 0 \tag{134}$$

ist.

Nach einem bekannten Satze der Elementarteilertheorie gibt es eine Matrix $\mathfrak{H} = (h_{kl})$, sodass die Matrix

$$\mathfrak{H}^{-1}\mathfrak{A}\mathfrak{H} = \mathfrak{L} \tag{135}$$

die Normalform besitzt. Zunächst werde angenommen, dass die Wurzeln λ_k $(k = 1, \cdots, n)$ sämtlich verschieden sind. Dann ist $\mathfrak{L}$ die Diagonalmatrix aus den Diagonalelementen $\lambda_1, \cdots, \lambda_n$. Man kann noch voraussetzen, dass für je zwei konjugiert komplexe Wurzeln λ_j und λ_k auch die j^{te} und die k^{te} Spalte von $\mathfrak{H}$ zueinander konjugiert komplex sind. Macht man die Substitution

$$\delta_k = \sum_{l=1}^{n} h_{kl}\zeta_l \qquad (k = 1, \cdots, n), \tag{136}$$

so sind $\delta_1, \cdots, \delta_n$ dann und nur dann reell, wenn für jedes Paar konjugiert komplexer Wurzeln λ_j, $\lambda_k = \bar{\lambda}_j$ auch stets $\zeta_k = \bar{\zeta}_j$ gilt und für reelles λ_k auch ζ_k reell ist. Wir wollen weiterhin nur solche Werte der Variabeln ζ_k betrachten, die diesen Bedingungen genügen.

Durch die lineare Substitution (136) geht das System (133) wegen (135) über in

$$\frac{d\zeta_k}{ds} = \lambda_k\zeta_k + \chi_k \qquad (k = 1, \cdots, n); \tag{137}$$

dabei sind die $\chi_k = \chi_k(\zeta_1, \cdots, \zeta_n)$ Potenzreihen in $\zeta_1, \cdots, \zeta_n$ ohne konstante und lineare Glieder, welche den Gleichungen

$$\psi_k = \sum_{l=1}^{n} h_{kl}\chi_l \qquad (k = 1, \cdots, n)$$

genügen. Für $\lambda_k = \bar{\lambda}_j$ ist also auch $\chi_k = \bar{\chi}_j$.

Wir wollen das System (137) weiter vereinfachen durch Substitutionen der Gestalt

(138) $$u_k = \zeta_k - P_k(\zeta_1, \cdots, \zeta_p) \qquad (k = 1, \cdots, n),$$

wo die P_k Potenzreihen der ersten p Variabeln $\zeta_1, \cdots, \zeta_p$ ohne konstante und lineare Glieder bedeuten. Wir setzen diese Potenzreihen zunächst mit unbestimmten Koeffizienten an und wollen diese dann rekursiv eindeutig festlegen durch gewisse Bedingungen. Die Untersuchung der Konvergenz werden wir nachträglich durchführen. Führt man die Variabeln u_k in die Differentialgleichungen (137) ein, so erhält man das System

(139) $$\frac{du_k}{ds} = \lambda_k u_k + Q_k(u_1, \cdots, u_n) \qquad (k = 1, \cdots, n)$$

mit

$$Q_k = \chi_k + \lambda_k P_k - \sum_{l=1}^{p} \frac{\partial P_k}{\partial \zeta_l}(\chi_l + \lambda_l \zeta_l),$$

wobei in χ_k, P_k, $\dfrac{\partial P_k}{\partial \zeta_l}$, χ_l, ζ_l die Variabeln $u_1, \cdots, u_n$ durch Auflösung von (138) nach $\zeta_1, \cdots, \zeta_n$ einzutragen sind. Die Q_k sind dann offenbar Potenzreihen in $u_1, \cdots, u_n$ ohne konstante und lineare Glieder. Nun wollen wir die Koeffizienten der P_k so zu bestimmen versuchen, dass in keiner der Reihen Q_k Potenzprodukte der p Variabeln $u_1, \cdots, u_p$ allein auftreten; es soll also identisch

(140) $$Q_k(u_1, \cdots, u_p, 0, \cdots, 0) = 0 \qquad (k = 1, \cdots, n)$$

gelten.

Nach (138) sind $\zeta_1, \cdots, \zeta_p$ Potenzreihen der p Variabeln $u_1, \cdots, u_p$ allein, und für $u_{p+1} = 0, \cdots, u_n = 0$ ist ausserdem

$$\zeta_k = P_k(\zeta_1, \cdots, \zeta_p) \qquad (k = p + 1, \cdots, n).$$

Daher sind die Bedingungen (140) gleichbedeutend mit

(141) $$\begin{aligned}&\chi_k(\zeta_1, \cdots, \zeta_p, P_{p+1}, \cdots, P_n) + \lambda_k P_k \\ &\qquad - \sum_{l=1}^{p} \frac{\partial P_k}{\partial \zeta_l}\{\chi_l(\zeta_1, \cdots, \zeta_p, P_{p+1}, \cdots, P_n) + \lambda_l \zeta_l\} = 0\end{aligned}$$

für $k = 1, \cdots, n$, identisch in $\zeta_1, \cdots, \zeta_p$. Wir zerlegen

$$P_k = \sum_{q=2}^{\infty} P_{kq} \qquad (k = 1, \cdots, n),$$

wo P_{kq} ein homogenes Polynom q^{ten} Grades in $\zeta_1, \cdots, \zeta_p$ bedeutet; entsprechend sei R_{kq} der homogene Bestandteil q^{ten} Grades in $\zeta_1, \cdots, \zeta_p$ auf der linken Seite

von (141). Dann gilt offenbar

$$R_{kq} = \lambda_k P_{kq} - \sum_{l=1}^{p} \frac{\partial P_{kq}}{\partial \zeta_l} \lambda_l \zeta_l + H_{kq}, \tag{142}$$

wo H_{kq} ein Polynom in $\zeta_1, \cdots, \zeta_p, P_{lh}$ $(l = p + 1, \cdots, n; h = 2, \cdots, q - 1)$ und $\dfrac{\partial P_{kh}}{\partial \zeta_l}$ $(l = 1, \cdots, p; h = 2, \cdots, q - 1)$ ist. Es seien bereits alle P_{lh} für $l = 1, \cdots, n$ und $h = 2, \cdots, q - 1$ bekannt; diese Annahme ist inhaltslos für $q = 2$. Dann haben wir P_{kq} $(k = 1, \cdots, n)$ zufolge (141), (142) aus den Bedingungen

$$-\lambda_k P_{kq} + \sum_{l=1}^{p} \frac{\partial P_{kq}}{\partial \zeta_l} \lambda_l \zeta_l = H_{kq} \tag{143}$$

zu ermitteln, deren rechte Seiten bekannt sind.

Ist $c\zeta_1^{g_1} \cdots \zeta_p^{g_p}$ ein Summand von H_{kq}, so ergibt sich für den Koeffizienten σ des entsprechenden Gliedes von P_{kq} aus (143) die lineare Gleichung

$$\left(-\lambda_k + \sum_{l=1}^{p} g_l \lambda_l\right)\sigma = c,$$

und danach erhält man σ, und zwar eindeutig, wenn

$$\sum_{l=1}^{p} g_l \lambda_l \neq \lambda_k \tag{144}$$

ist. Da $g_1, \cdots, g_p$ nicht-negative ganze Zahlen mit der Summe $q \geqq 2$ und die reellen Teile von $\lambda_1, \cdots, \lambda_p$ negativ, die von $\lambda_{p+1}, \cdots, \lambda_n$ positiv sind, so ist (144) für $k > p$ stets erfüllt und für $k = 1, \cdots, p$ jedenfalls bei genügend grossem q. Wir setzen zunächst voraus, dass (144) ausnahmslos erfüllt ist, für $k = 1, \cdots, n$ und alle Systeme nicht-negativer ganzer Zahlen $g_1, \cdots, g_p$, deren Summe grösser als 1 ist. Dann sind also die Potenzreihen P_k auf genau eine Weise so bestimmbar, dass (140) erfüllt wird.

Es muss nun untersucht werden, wie es mit der Konvergenz der formal gefundenen Reihen P_k bestellt ist. Diese Untersuchung erfolgt am bequemsten mit der Cauchyschen Majorantenmethode. Sind P und Q zwei Potenzreihen in den gleichen Variabeln, so soll das Zeichen

$$P \prec Q$$

bedeuten, dass für je zwei entsprechende Koeffizienten α und β von P und Q die Ungleichung

$$|\alpha| \leqq \beta$$

gilt. Nach Voraussetzung sind die Potenzreihen ψ_k in (133), also auch die Reihen χ_k in (137) konvergent in einer gewissen Umgebung des Nullpunktes. Bedeutet C_1, wie auch weiterhin $C_2, \cdots, C_6$ eine geeignete positive Konstante,

so ist also

$$\chi_k \prec \frac{C_1(\zeta_1 + \cdots + \zeta_n)^2}{1 - C_1(\zeta_1 + \cdots + \zeta_n)} = f(\zeta_1, \cdots, \zeta_n) \quad (k = 1, \cdots, n),$$

wo die rechte Seite durch ihre Potenzreihe zu ersetzen ist.

Wegen (144) ist nun

$$(145) \qquad 1 + g_1 + \cdots + g_p < C_2 \left| -\lambda_k + \sum_{l=1}^{p} g_l \lambda_l \right|,$$

für $k = 1, \cdots, n$ und jedes System nicht-negativer ganzer Zahlen $g_1, \cdots, g_p$, deren Summe mindestens gleich 2 ist. Wir ersetzen (141) durch die abgeänderten Gleichungen

$$(146) \qquad P_k^* + \sum_{l=1}^{p} \frac{\partial P_k^*}{\partial \zeta_l} \zeta_l = C_2 \left(1 + \sum_{l=1}^{p} \frac{\partial P_k^*}{\partial \zeta_l}\right) f(\zeta_1, \cdots, \zeta_p, P_{p+1}^*, \cdots, P_n^*)$$

$$(k = 1, \cdots, n)$$

für n unbekannte Potenzreihen $P_k^*(\zeta_1, \cdots, \zeta_p)$. Diese lassen sich durch dasselbe rekursive Verfahren bestimmen wie die P_k, und man entnimmt durch Vergleich von (141) und (146) wegen (145) unmittelbar die Beziehung

$$(147) \qquad P_k \prec P_k^* \qquad (k = 1, \cdots, n).$$

Aus (146) ersieht man, dass alle P_k^* $(k = 1, \cdots, n)$ identisch gleich sind; es sei P^* der gemeinsame Wert. Setzt man noch $\zeta_1 = \zeta_2 = \cdots = \zeta_p = \zeta$, so möge P^* in $P(\zeta)$ übergehen; ferner sei

$$f(\zeta, P) = \frac{C_1\{p\zeta + (n - p)P\}^2}{1 - C_1\{p\zeta + (n - p)P\}}.$$

Offenbar ist dann

$$(148) \qquad P^*(\zeta_1, \cdots, \zeta_p) \prec P(\zeta_1 + \cdots + \zeta_p)$$

und nach (146)

$$P + \zeta \frac{dP}{d\zeta} = C_2 \left(1 + \frac{dP}{d\zeta}\right) f(\zeta, P).$$

Durch Einsetzen der Potenzreihe

$$P(\zeta) = \sum_{q=2}^{\infty} d_q \zeta^q$$

erhält man

$$\sum_{q=2}^{\infty} (q + 1)\, d_q \zeta^q = C_2 \left(1 + \sum_{q=2}^{\infty} q\, d_q \zeta^{q-1}\right) f(\zeta, P).$$

Die hieraus entstehenden Rekursionsformeln für die Koeffizienten d_q zeigen,

dass

$$P \prec \zeta M_1 \tag{149}$$

ist, wo M_1 der kubischen Gleichung

$$\zeta M_1 = \frac{C_3(\zeta + \zeta M_1)^2}{1 - C_3(\zeta + \zeta M_1)} (1 + M_1)$$

genügt. Nun ist

$$\frac{(1 + M_1)^3}{1 - C_3\zeta(1 + M_1)} \prec \frac{1}{1 - C_4(\zeta + M_1)}, \qquad C_3\zeta \prec M_1$$

und folglich

$$M_1 \prec M_2 \tag{150}$$

mit

$$M_2 = \frac{C_5\zeta}{1 - C_5 M_2}.$$

Aus der Gleichung

$$(1 - 2C_5M_2)^{-2} = (1 - 4C_5^2\zeta)^{-1}$$

erhält man endlich

$$M_2 \prec M \tag{151}$$

für

$$1 + 4C_5M = (1 - 4C_4^2\zeta)^{-1},$$
$$M = C_5\zeta(1 - 4C_5^2\zeta)^{-1},$$

also nach (147), (148), (149), (150), (151)

$$P_k \prec \frac{C_6(\zeta_1 + \cdots + \zeta_p)^2}{1 - C_6(\zeta_1 + \cdots + \zeta_p)} \qquad (k = 1, \cdots, n).$$

Damit ist die Konvergenz der Potenzreihen P_k für hinreichend kleine Werte der absoluten Beträge von $\zeta_1, \cdots, \zeta_p$ bewiesen.

Schliesslich sind noch die Realitätsverhältnisse zu untersuchen. Aus der Lösung $P_1, \cdots, P_n$ von (141) entsteht eine weitere Lösung, indem man für jede reelle Wurzel λ_k die Grösse P_k durch $\bar{P}_k$ ersetzt und für je zwei konjugiert komplexe Wurzeln λ_j, λ_k das Paar P_j, P_k durch $\bar{P}_k, \bar{P}_j$. Da aber die Lösung eindeutig festgelegt ist, so gilt $P_k = \bar{P}_k$ im ersten Falle und $P_k = \bar{P}_j$ im zweiten Falle. Nach (138) ist also u_k reell für reelles λ_k und $u_k = \bar{u}_j$ für $\lambda_k = \bar{\lambda}_j$.

Wir wollen nunmehr sämtliche Lösungen der Differentialgleichungen (133) bestimmen, die für $s \to \infty$ in den Nullpunkt einmünden. Nach (136) und (138) genügt es, diese Untersuchung an dem System (139) vorzunehmen. Da s nicht explizit in den Differentialgleichungen auftritt, so kann man voraussetzen, dass

die gesuchten Lösungen für alle $s \geqq 0$ der Ungleichung

$$\sum_{k=1}^{n} |u_k|^2 < \epsilon \tag{152}$$

genügen, wo ϵ eine hinreichend klein zu wählende positive Konstante bedeutet. Setzt man

$$\sum_{k=p+1}^{n} |u_k|^2 = Z, \tag{153}$$

so gilt zufolge (139)

$$\frac{dZ}{ds} = \sum_{k=p+1}^{n} (\lambda_k + \bar{\lambda}_k)u_k\bar{u}_k + \sum_{k=p+1}^{n} (u_k\bar{Q}_k + \bar{u}_kQ_k). \tag{154}$$

Nach (140) ist nun jedes Glied der Potenzreihen Q_k und $\bar{Q}_k$ durch eine der Variabeln $u_{p+1}, \cdots, u_n$ teilbar, also jedes Glied der Potenzreihe für den zweiten Summanden auf der rechten Seite von (154) durch ein Produkt zweier dieser Variabeln teilbar. Da diese Potenzreihe mit Gliedern dritter Ordnung beginnt, so ist nach (152) und (153) für genügend kleines ϵ die Ungleichung

$$\sum_{k=p+1}^{n} (u_k\bar{Q}_k + \bar{u}_kQ_k) \geqq -\rho_n Z \tag{155}$$

erfüllt, also nach (134), (153), (154)

$$\frac{dZ}{ds} \geqq 2 \sum_{k=p+1}^{n} \rho_k |u_k|^2 - \rho_n Z \geqq \rho_n Z,$$

$$\frac{d(Ze^{-\rho_n s})}{ds} \geqq 0.$$

Daher ist der Ausdruckt $Ze^{-\rho_n s}$ für alle $s \geqq 0$ monoton wachsend; andererseits strebt er für $s \rightarrow \infty$ nach 0, weil ρ_n positiv und $Z < \epsilon$ ist. Folglich ist für die gesuchte Lösung $Z = 0$, also

$$u_k = 0 \qquad (k = p + 1, \cdots, n). \tag{156}$$

Nach (140) reduzieren sich jetzt die Differentialgleichungen (139) auf die einfache Form

$$\frac{du_k}{ds} = \lambda_k u_k \qquad (k = 1, \cdots, p) \tag{157}$$

mit der Lösung

$$u_k = \alpha_k e^{\lambda_k s} \qquad (k = 1, \cdots, p). \tag{158}$$

Dabei ist α_k reell für reelles λ_k und $\alpha_k = \bar{\alpha}_j$ für $\lambda_k = \bar{\lambda}_j$. Da die Realteile von $\lambda_1, \cdots, \lambda_p$ negativ sind, so mündet die durch (156) und (158) gegebene Lösung bei beliebigen α_k tatsächlich für $s \rightarrow \infty$ in den Nullpunkt ein.

Durch die Bedingungen (156) wird eine analytische Fläche von p Dimensionen im Raume der $\delta_1, \cdots, \delta_n$ definiert. Man erhält eine Darstellung der Fläche durch die Parameter $\zeta_1, \cdots, \zeta_p$, wenn man in die Ausdrücke

$$\delta_k = \sum_{l=1}^{n} h_{kl}\zeta_l \qquad (k = 1, \cdots, n)$$

für $\zeta_{p+1}, \cdots, \zeta_n$ die aus (138), (156) folgenden Werte

$$\zeta_k = P_k(\zeta_1, \cdots, \zeta_p) \qquad (k = p + 1, \cdots, n)$$

einträgt. Will man eine reelle Parameterdarstellung haben, so hat man das Paar ζ_j, ζ_k für $\lambda_k = \bar{\lambda}_j$ durch $(\zeta_j + \zeta_k)/2$, $(\zeta_j - \zeta_k)/2i$ zu ersetzen. Die gesuchten Bahnkurven erfüllen dann genau diese p-dimensionale Fläche, und zwar erhält man die einzelnen Lösungen, indem man vermöge der Gleichungen

$$u_k = \zeta_k - P_k(\zeta_1, \cdots, \zeta_p) \qquad (k = 1, \cdots, p)$$

die Grössen $\zeta_1, \cdots, \zeta_p$ als Potenzreihen von $u_1, \cdots, u_p$ bestimmt und dann

$$u_k = \alpha_k e^{\lambda_k s}$$

setzt, mit beliebigen Konstanten α_k $(k = 1, \cdots, p)$ unter Beachtung der Realitätsbedingungen. Die allgemeine Lösung hängt also von p reellen Parametern ab.

Es seien noch einmal die beiden Voraussetzungen erwähnt, die wir in diesem Paragraphen an verschiedenen Stellen der Untersuchung eingeführt haben, nämlich die Verschiedenheit der Wurzeln $\lambda_1, \cdots, \lambda_n$ und das Bestehen der Ungleichung (144). Wir wollen noch feststellen, in welcher Art unsere Ergebnisse zu modifizieren sind, wenn wir diese Voraussetzungen fallen lassen.

12. Ausartungen

Wir verzichten nun auf die Annahme (144), halten aber zunächst noch an der Voraussetzung fest, dass $\lambda_1, \cdots, \lambda_n$ verschieden sind. Wir denken uns dann die endlich vielen Lösungssysteme der diophantischen Gleichung

$$\sum_{l=1}^{p} g_l\lambda_l = \lambda_k \tag{159}$$

bestimmt, wobei k irgend ein Index der Reihe $1, \cdots, p$ ist und $g_1, \cdots, g_p$ nicht-negative ganze Zahlen, deren Summe mindestens 2 beträgt. In dem Ansatz (138) waren bisher $P_1, \cdots, P_n$ Potenzreihen in $\zeta_1, \cdots, \zeta_p$ ohne konstante und lineare Glieder, deren Koeffizienten rekursiv eindeutig durch die Bedingung (140) festgelegt wurden. Jetzt schliessen wir aus dem mit unbestimmten Koeffizienten gebildeten Potenzreihen $P_1, \cdots, P_n$ sämtliche Glieder $\sigma\zeta_1^{g_1} \cdots \zeta_p^{g_p}$ aus, deren Exponenten einer Gleichung (159) genügen, und ersetzen (140) durch die folgende abgeschwächte Bedingung: Für $k = 1, \cdots, p$ soll eine Identität

$$Q_k(u_1, \cdots, u_p, 0, \cdots, 0) = V_k(u_1, \cdots, u_p) \tag{160}$$

gelten, wo V_k ein Polynom aus solchen Gliedern $cu_1^{g_1} \cdots u_p^{g_p}$ bedeutet, deren

Exponenten (159) erfüllen; für $k = p + 1, \cdots, n$ soll wie bisher

$$Q_k(u_1, \cdots, u_p, 0, \cdots, 0) = 0 \tag{161}$$

gelten. Es lässt sich auf dem früheren Wege ohne Mühe zeigen, dass die Potenzreihen $P_1, \cdots, P_n$ wieder eindeutig bestimmt und konvergent sind, und auch die Polynome $V_1, \cdots, V_p$ sind eindeutig fixiert. Da zur Herleitung der Ungleichung (155) die Formeln (140) nur für $k = p + 1, \cdots, n$ herangezogen werden, so bleibt die zu (156) führende Schlussweise bestehen. Bei sämtlichen für $s \to \infty$ nach dem Nullpunkt wandernden Lösungen von (133) ist also wieder $u_{p+1} = 0, \cdots, u_n = 0$. An die Stelle des Systems (157) tritt aber jetzt nach (139) und (160)

$$\frac{du_k}{ds} = \lambda_k u_k + V_k(u_1, \cdots, u_p) \qquad (k = 1, \cdots, p). \tag{162}$$

Nach (134) ist für jede Lösung von (159)

$$g_k = 0, \qquad g_{k+1} = 0, \cdots, g_p = 0,$$

folglich ist $V_k = V_k(u_1, \cdots, u_{k-1})$ ein Polynom in $u_1, \cdots, u_{k-1}$ allein und $V_1 = 0$. Die Integration von (162) lässt sich ohne weiteres ausführen und ergibt

$$u_1 = \alpha_1 e^{\lambda_1 s},$$

$$u_2 = \alpha_2 e^{\lambda_2 s} + sV_2(u_1) = \{\alpha_2 + sV_2(\alpha_1)\} e^{\lambda_2 s}, \tag{163}$$

allgemein durch vollständige Induktion

$$u_k = (\alpha_k + W_k) e^{\lambda_k s} \qquad (k = 1, \cdots, p), \tag{164}$$

wo W_k ein eindeutig bestimmtes Polynom in $\alpha_1, \cdots, \alpha_{k-1}$ und s ist. Da umgekehrt bei beliebiger Wahl der Konstanten $\alpha_1, \cdots, \alpha_p$ die durch (164) gegebenen Funktionen u_k für $s \to \infty$ gegen 0 streben, so gelten die Ergebnisse des vorangehenden Paragraphen auch in diesem Fall, wenn nur u_k durch (164) erklärt wird.

Endlich lassen wir auch noch die Voraussetzung fallen, dass die Wurzeln $\lambda_1, \cdots, \lambda_n$ alle verschieden sind. Dann ist die Matrix $\mathfrak{C} = (q_{kl})$ in (135) im allgemeinen keine reine Diagonalmatrix: Es ist wieder $q_{kk} = \lambda_k$, $q_{kl} = 0$ für $k \neq l$ und $k \neq l + 1$, $q_{kl} = 0$ für $k = l + 1$ und $\lambda_l \neq \lambda_{l+1}$; für $k = l + 1$ und $\lambda_l = \lambda_{l+1}$ ist dagegen q_{kl} entweder gleich 0 oder gleich λ_l. An die Stelle des Systems (137) tritt jetzt allgemeiner

$$\frac{d\zeta_1}{ds} = \lambda_1 \zeta_1 + \chi_1,$$

$$\frac{d\zeta_k}{ds} = \lambda_k(\zeta_k + e_k \zeta_{k-1}) + \chi_k \qquad (k = 2, \cdots, n);$$

dabei ist $e_k = 0$ für $\lambda_k \neq \lambda_{k-1}$, während im Falle $\lambda_k = \lambda_{k-1}$ entweder $e_k = 0$ oder $e_k = 1$ ist. Anstatt (135) erhält man durch die früher benutzte Methode

die Differentialgleichungen

$$(165)\qquad \begin{cases} \dfrac{du_1}{ds} = \lambda_1 u_1 + Q_1(u_1, \cdots, u_n), \\ \dfrac{du_k}{ds} = \lambda_k(u_k + e_k u_{k-1}) + Q_k(u_1, \cdots, u_n) \qquad (k = 2, \cdots, n), \end{cases}$$

wo $Q_1, \cdots, Q_n$ wieder den Bedingungen (160) und (161) genügen, mit der dort angegebenen Bedeutung der Polynome $V_1, \cdots, V_p$. Anstelle des in (153) erklärten Ausdrucks Z betrachten wir allgemeiner

$$\sum_{k=p+1}^{n} h_k \,|\, u_k \,|^2 = Z_0$$

mit konstanten Werten $h_{p+1}, \cdots, h_n$. Zufolge (165) gilt dann

$$(166)\qquad \begin{aligned} \frac{dZ_0}{ds} = \sum_{k=p+1}^{n} h_k\{(\lambda_k + \bar{\lambda}_k) \,|\, u_k \,|^2 + e_k(\lambda_k u_{k-1} \bar{u}_k + \bar{\lambda}_k \bar{u}_{k-1} u_k)\} \\ + \sum_{k=p+1}^{n} (u_k \bar{Q}_k + \bar{u}_k Q_k). \end{aligned}$$

Wegen $\lambda_p \neq \lambda_{p+1}$ ist nun $e_{p+1} = 0$, und der erste Summand auf der rechten Seite von (166) lässt sich in der Form

$$\begin{aligned} H = \rho_n h_n \,|\, u_n \,|^2 + \sum_{k=p+2}^{n} \left(\rho_{k-1} h_{k-1} - e_k \frac{|\,\lambda_k\,|^2}{\rho_k} h_k\right) |\, u_{k-1} \,|^2 \\ + \sum_{k=p+1}^{n} \rho_k h_k \left| u_k + e_k \frac{\lambda_k}{\rho_k} u_{k-1} \right|^2 \end{aligned}$$

schreiben. Wählt man speziell

$$h_{p+1} = 1, \qquad h_{k+1} = \frac{\rho_k \rho_{k+1}}{2\,|\,\lambda_{k+1}\,|^2} h_k \qquad (k = p + 1, \cdots, n - 1),$$

so wird offenbar

$$(167)\qquad H \geqq \tfrac{1}{2} \sum_{k=p+1}^{n} \rho_k h_k \,|\, u_k \,|^2 \geqq \tfrac{1}{2} \rho_n Z_0.$$

Aus (166) und (167) folgt jetzt, wenn ϵ in (152) genügend klein gewählt wird, die Ungleichung

$$\frac{dZ_0}{ds} \geqq \tfrac{1}{3} \rho_n Z_0$$

und daraus wieder das Verschwinden von $u_{p+1}, \cdots, u_n$ bei sämtlichen Lösungen von (165), die für $s \rightarrow \infty$ im Nullpunkt einmünden. Es bleibt noch das System

$$\frac{du_1}{ds} = \lambda_1 u_1,$$

$$\frac{du_k}{ds} = \lambda_k(u_k + e_k u_{k-1}) + V_k(u_1, \cdots, u_{k-1}) \qquad (k = 2, \cdots, p)$$

zu integrieren. Die Lösung hat wieder die Gestalt

$$u_k = (\alpha_k + W_k)e^{\lambda_k s} \qquad (k = 1, \cdots, p),$$

wo W_k ein eindeutig bestimmtes Polynom in s und den Integrationskonstanten $\alpha_1, \cdots, \alpha_{k-1}$ bedeutet; insbesondere ist

$$W_1 = 0, \qquad W_2 = s\{c_2\lambda_2\alpha_1 + V_2(\alpha_1)\}.$$

Damit haben wir die Resultate des vorigen Paragraphen auf den Fall mehrfacher Wurzeln $\lambda_1, \cdots, \lambda_n$ übertragen, abgesehen von der Untersuchung der Realitätsverhältnisse. Diese Untersuchung liesse sich ohne Schwierigkeit durchführen; wir gehen aber darauf nicht mehr ein, da bei der Anwendung auf die Dreierstosslösungen im Falle einer mehrfachen Wurzel alle Wurzeln reell sind und dann überhaupt keine imaginären Grössen in die Rechnung eingeführt zu werden brauchen.

13. Die Dreierstossbahnen

Wir wenden nunmehr die Ergebnisse der beiden vorangehenden Paragraphen auf die Untersuchung des Systemes (118) an. Die charakteristischen Wurzeln der Matrix $\mathfrak{A}$ werden im gleichseitigen Falle durch die 6 Zahlen $-a_0$, $-a_1$, $-a_2$, a_3, a_4, a_5 in (131) geliefert, im geradlinigen Falle durch die Zahlen $-b_0$, $-b_1$, b_2, b_3, b_4, b_5 in (132). Es sind $-a_0$, $-a_1$, $-a_2$, $-b_0$, $-b_1$ negativ und die übrigen Wurzeln sind positiv oder haben positiven Realteil. Folglich ist $p = 3$ im gleichseitigen Falle und $p = 2$ im geradlinigen Falle.

Die Wurzeln $-a_0$, $-a_1$, $-a_2$ sind sämtlich verschieden, ausser wenn $m_1 = m_2 = m_3$ ist, und dann ist $-a_1 = -a_2 = (1 - \sqrt{13})/6 > -a_0$. Die Wurzeln $-b_0$, $-b_1$ sind stets voneinander verschieden.

Es ist noch festzustellen, wann (159) lösbar ist und welches dann die Lösungen sind. Im gleichseitigen Falle ist

$$a_0 > a_1 \geqq a_2, \qquad 2a_1 > a_0, \qquad a_1 + a_2 > a_0;$$

gilt also

$$a_0x + a_1y + a_2z = a_k, \qquad x + y + z \geqq 2$$

für ein k der Reihe 1, 2, 3 in nicht-negativen ganzen Zahlen x, y, z, so folgt $x = 0$, $y = 0$, und es ist entweder

(168) $$a_0 = ga_2, \qquad z = g$$

mit ganzem g oder

(169) $$a_1 = ha_2, \qquad z = h$$

mit ganzem h. Die Annahme (168) ergibt nach (131) für die in (128) definierte Grösse

$$a = \frac{m_2m_3 + m_3m_1 + m_1m_2}{(m_1 + m_2 + m_3)^2}$$

den Wert

$$a = \tfrac{4}{27} g^{-4}(g+2)(3g^2 - g - 2) \qquad (g = 2, 3, \cdots), \tag{170}$$

während aus (169)

$$a = \tfrac{4}{3} h(h^2+1)^{-2}(hv+1)(h-v) \qquad (h = 1, 2, \cdots) \tag{171}$$

mit

$$v = \tfrac{1}{12}(h^2-1)(h^2+1)^{-1}\{-1 + [1 + 24(h^2+1)(h+1)^{-2}]^{\frac{1}{2}}\}$$

folgt. Hieraus ist noch leicht ersichtlich, dass nicht (168) und (169) zugleich eintreten können. Beim geradlinigen Falle ist $b_1 > b_0$; gilt also

$$b_0 x + b_1 y = b_k , \qquad x + y \geqq 2$$

für ein k der Reihe 1, 2 in nicht-negativen ganzen Zahlen x, y, so ist

$$b_1 = j b_0 , \qquad x = j, \qquad y = 0$$

mit ganzem j, und die in (130) erklärte Grösse

$$b = \frac{m_1\{1 + (1-\omega)^{-1} + (1-\omega)^{-2}\} + m_3(1 + \omega^{-1} + \omega^{-2})}{m_1 + m_2\{\omega^{-2} + (1-\omega)^{-2}\} + m_3}$$

hat nach (132) den Wert

$$b = j^2 + \tfrac{1}{2}(j - 3). \tag{172}$$

Im gleichseitigen Falle bilden die für $s \to \infty$ in den Nullpunkt einmündenden Lösungen von (118) eine dreidimensionale analytische Mannigfaltigkeit. Ist nicht (170) oder (171) erfüllt, so lassen sich $\delta_1 , \cdots , \delta_6$ in Potenzreihen der Variabeln

$$u_1 = \alpha_1 e^{-a_0 s}, \qquad u_2 = \alpha_2 e^{-a_1 s}, \qquad u_3 = \alpha_3 e^{-a_2 s} \tag{173}$$

entwickeln; dabei hängen die Koeffizienten der Potenzreihen nur von m_1 , m_2 , m_3 ab und α_1 , α_2 , α_3 sind willkürliche reelle Konstanten. Ist (170) erfüllt, so hat man auf Grund von (163)

$$u_1 = (\alpha_1 - c_1 \alpha_3^g s) e^{-a_0 s} \tag{174}$$

zu setzen, wo c_1 eine von m_1 , m_2 , m_3 abhängige Konstante bedeutet. Im Falle (171) ist entsprechend

$$u_2 = (\alpha_2 - c_2 \alpha_3^h s) e^{-a_1 s}. \tag{175}$$

Im geradlinigen Falle bilden die für $s \to \infty$ in den Nullpunkt einmündenden Lösungen von (118) eine zweidimensionale analytische Mannigfaltigkeit. Ist nicht (172) erfüllt, so lassen sich $\delta_1 , \cdots , \delta_6$ in Potenzreihen der Variabeln

$$v_1 = \beta_1 e^{-b_0 s}, \qquad v_2 = \beta_2 e^{-b_1 s}$$

entwickeln, wo die Koeffizienten nur von m_1 , m_2 , m_3 abhängen und β_1 , β_2

willkürliche reelle Konstanten bedeuten. Ist (172) erfüllt, so hat man

$$v_2 = (\beta_2 - c_3\beta_1^j s)e^{-b_1 s}$$

zu setzen.

Trägt man die gefundenen Potenzreihen für $\delta_1, \cdots, \delta_6$ in die rechte Seite von (119) ein, so erhält man $\delta_8 = p_4$ durch eine einfache Quadratur. Da das unbestimmte Integral einer Potenzreihe in den durch (173) oder (174), (175) definierten Grössen u_1, u_2, u_3 wieder eine solche Potenzreihe ist und das gleiche für die Potenzreihen in v_1, v_2 gilt, so ist also

$$p_4 = \gamma + P,$$

wo γ eine willkürliche Konstante und P eine eindeutig festgelegte Potenzreihe in u_1, u_2, u_3 oder in v_1, v_2 ohne konstantes Glied ist. Folglich hat p_4 einen Grenzwert für $s \to \infty$, nämlich die Zahl γ. Damit ist also endlich bewiesen, dass die 3 Massenpunkte im ruhenden Koordinatensystem in bestimmten Richtungen zusammenstossen. Durch eine geeignete Drehung des Koordinatensystems kann man erreichen, dass $\gamma = 0$ ist.

Trägt man die gefundenen Reihen nach (96), (109), (114) in die Werte der Koordinaten x_k, y_k ($k = 1, 2, 3$) ein und benutzt (110), so erhält man im gleichseitigen Falle für jede Koordinate eine Darstellung

$$w = t^{2/3}w^*(u_1, u_2, u_3),$$

wo $w^*(u_1, u_2, u_3)$ eine Reihe nach positiven Potenzen von

$$u_1 = \alpha_1 t^{2/3}, \qquad u_2 = \alpha_2 t^{a_1}, \qquad u_3 = \alpha_3 t^{a_2}$$

bedeutet, deren Koeffizienten nur von m_1, m_2, m_3 abhängen. Liegt eine der Ausartungen (170) oder (171) vor, so ist statt dessen

$$u_1 = (\alpha_1 + c_1\alpha_3^g \log t)t^{2/3}$$

oder

$$u_2 = (\alpha_2 + c_2\alpha_3^h \log t)t^{a_1}$$

zu setzen. Im geradlinigen Falle hat man analog

$$w = t^{2/3}w^*(v_1, v_2)$$

mit

$$v_1 = \beta_1 t^{2/3}, \qquad v_2 = \beta_2 t^{b_1}$$

oder, wenn (172) erfüllt ist, mit

$$v_2 = (\beta_2 + c_3\beta_1^j \log t)t^{b_1}.$$

Schliesslich kann man noch eine Drehung des Koordinatensystems um konstante Winkel vornehmen und ihm eine Translation mit konstanter Geschwindigkeit erteilen.

Die Werte a_1, a_2 und b_1 sind irrational, wenn m_1, m_2, m_3 nicht im Bereich der rationalen Zahlen in gewisser Weise algebraisch abhängig voneinander sind. Daher ist der Punkt $t = 0$ im allgemeinen ein logarithmischer Verzweigungspunkt für die Koordinaten der Dreierstossbahnen.

[Zusatz bei der Korrektur:] Inzwischen bin ich aufmerksam geworden auf eine Arbeit von G. Sokoloff,[6] welche ebenfalls der Untersuchung des Dreierstosses gewidmet ist und Reihenentwicklungen für die Koordinaten der kollidierenden Massenpunkte enthält. Im Beweis findet sich eine Lücke, die aber durch Benutzung des Bohlschen Satzes leicht ausgefüllt werden kann. Die vorliegende Darstellung dürfte in wesentlichen Punkten den Vorzug grösser Einfachheit haben.

PRINCETON, N. J.

[6] G. Sokoloff, *Conditions d'une collision générale des trois corps qui s'attirent mutuellement suivant la loi de Newton*, Académie des Sciences de l'Ukraine, Mémoires de la classe des sciences physiques et mathématiques, Bd. 9 (1928), S. 1–64.

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

INTUITIVE PROBABILITIES AND SEQUENCES[1]

BY B. O. KOOPMAN

(Received March 22, 1940)

This paper forms a second part of the general study of probability regarded as a branch of intuitive logic, the first part having appeared in *The Axioms and Algebra of Intuitive Probability.*[2] The point of view, notation, assumptions and theorems of the latter are assumed in the present work. The chief object here is to set forth the connection between the objective notion of statistical weight or frequency in a sequence and the intuitive conceptions of our theory.

It is necessary first to examine the class of definitions of probability based on the notion of frequency in a sequence (collectives); we shall reach the conclusion that *every application of such definitions of probability to experimental science implicitly presupposes the a priori intuitive conception of probability* (§1). Further progress in the exact formulation of the ideas requires the scrutiny of certain logical discriminations involved in the relation of *asserted* and *contemplated* propositions, a discrimination the ignoring of which leads to fallacies (§2). The notion of *symmetry* and the related one of *independence* in a set of trials is then studied (§3); the former replaces in the present theory of sequences the notion of *random* in the usual theory of collectives. Finally the theorems relating statistical weight with intuitive probability (numerical and otherwise) are established.

There is thus achieved a complete synthesis of the intuitive concept of probability with the objective notion of statistical weight which forms the basis of statistics and quantum mechanics,—and this without the assumption of a single general principle beyond those which have already been posited in our earlier work.

1. Probability as a Frequency

An event E (such as the tossing of a coin, the search for a photon with a sensitive plate, the observation of fatality following an inoculation, etc.) is regarded as having two possible outcomes, "success" (labeled 1) and "failure" (labeled 0). An infinite sequence of trials of E are made (conceptually) under essentially the same conditions, the results (1 or 0) forming the sequence (α): $(\alpha_1, \alpha_2, \cdots)$ $(\alpha_i = 0 \text{ or } 1)$. The theory of probability in collectives starts with the following assumptions.[3]

[1] Presented to the American Mathematical Society April 26, 1940.

[2] Annals of Math., Vol. 41, No. 2 (1940). Presented to the American Mathematical Society February 25, 1939. This paper will be referred to herein as AAP.

[3] Cf., e.g., R. von Mises, *Probability, Statistics and Truth* (London 1939). A. H. Copeland, *Consistency of the Conditions Determining Kollektivs*, Trans. Amer. Math. Soc., Vol. 42 (1937) p. 333, as well as the references contained therein.

First assumption. $\lim_{n\to\infty}(\alpha_1 + \cdots + \alpha_n)/n = w$.

Second assumption. If (α'):$(\alpha'_1, \alpha'_2, \cdots)$ *is a sub-sequence of* (α) *defined by any place-selection* s *of the class* S *then* $\lim_{n\to\infty}(\alpha'_1 + \cdots + \alpha'_n)/n = w$.

Here a place-selection is regarded as defined by a precise mathematical rule, and the class S as having been exactly delimited: we shall raise no question on this point.[4] Any sequence (α) for which these assumptions are made is called a *collective* (properly: *with respect to* S), and according to what shall be termed herein the collectivist definition of probability, *the number* w (naturally $0 \leqq w \leqq 1$) *is defined as the probability of success of* E (*in the collective* (α)).

Our object is not with the mathematical theory of collectives. We wish merely to ascertain whether the collectivist definition of probability can be applied to an experimental phenomenon without an auxiliary notion of an *a priori* subjective order reducing in essence to that of intuitive probability. And since we are dealing only with Assumption 1, our critique will apply to the whole class of theories which regard probability as merely a frequency in a sequence.

Let us suppose that I am given the information that (α) is a collective, that I know the definition of S, and also the value of w (e.g., let $w = 0.1$). Furthermore, let me excise from my brain all *a priori* intuitive notions of "likelihood", "degree of conviction", "reasonableness", etc., so that only my strict logical and observational powers remain. Just what conclusion can I derive from this knowledge concerning (α) and applying to the results of the totality of trials of E which I can make in my lifetime? Since I can only make a number N of trials obviously $<10^{10}$, and inasmuch as any set of values $(\alpha_1, \cdots, \alpha_N)$ ($\alpha_n = 0$ or 1) will be mathematically consistent with the assumptions regarding (α), it follows that I can draw absolutely no conclusion at all (in the only terms which I can understand, i.e., strict logic), so my information regarding (α) is totally irrelevant to my experience with E.

The situation is exactly the same for any individual whose name we can give. Even if it could be proved that some individual exists for whose set of trials of E surely $(\alpha_1 + \cdots + \alpha_N)/N = 0.1$, we could never inform him that he was this favored one, nor could he ever know it himself—until, indeed, his trials of E are all completed, in which case the information concerning (α) would become unnecessary. And an individual can be replaced by all society during a given historical era: The number N of total possible trials which it can make is much larger, but once the era has been named an upper limit of N is set and any values of $(\alpha_1, \cdots, \alpha_N)$ are consistent with the assumptions concerning (α). It may be said that a time will come when all observations made thereunto will give for $(\alpha_1 + \cdots + \alpha_N)/N$ a value close to w. But this does not remove the difficulty: The time will never come when it can be *proved* that the proper value of N has been reached.

A finite theory of collectives has been proposed in which the finite set

[4] For a discussion from the point of view of modern logic cf. A. Church, *On the Concept of a Random Sequence*, Bull. Amer. Math. Soc. Vol. 46 (1940) p. 130.

$(\alpha_1, \cdots, \alpha_M)$ replaces the infinite sequence $(\alpha_1, \alpha_2, \cdots)$ and such ratios as $(\alpha_1 + \cdots + \alpha_N)/N$ replace the corresponding limits.[5] Here M is regarded as enormous—"practically infinite"—so that the previous N is minute in comparison, and any values of $(\alpha_1, \cdots, \alpha_N)$ are, here also, consistent with any values of the above ratio within limits of experimental error: The earlier difficulty is encountered here as before. This will naturally be quite otherwise if N is of the same order of magnitude as M; but in this case (even assuming that the experimenter could ever know it) the theory will be superfluous. The matter can be put thus: The problem of sampling is insoluble—nay, unstable—in terms of strict logic divorced from all intuitive notions of probability.

Certain philosophers regard the theory of collectives as affording not a theory of knowledge but a pattern of behavior.[6] But it is always a question not of *any* behavior but of a *most favorable* behavior in a certain highly precise sense. And the proposition that the particular pattern of behavior derived from the theory of collectives is actually this most favorable one becomes subject to all the aforementioned difficulties.

Leaving collectives for the moment, we consider the great class of physical quantities which are defined as limits. It soon appears that here also the result of no measurement of the sort which can be carried out in the laboratory either determines or is determined by what is deemed the true value of such a quantity by strict mathematical logic alone.[7] For example the density ρ of a continuous medium at the point P is defined as $\rho = \lim M/V$ (as $V \to P$): The finite set of ratios M/V which I can ever measure have values totally devoid of strict logical relationship with the value of ρ. Nor is the difficulty due to the idealization involved in the definition of ρ: If ρ is defined simply as the value of M/V measured by a specified person who has instruments a hundred times more effective than any which I possess, the difficulty remains the same (cf. the case of finite collectives).

Of course, when I find that my values of M/V for rapidly decreasing V vary by amounts growing negligible, the *common sense* of the experimenter leads me to the conclusion that the final will not differ appreciably from the actual value of ρ. But precisely what does it mean to conclude something by common sense which is not derivable by strict logic? If we have made the meaning of intuitive probability in the least degree clear, it must be apparent that such a conclusion of common sense is precisely an instance of a conceptualism of intuitive probability. Thus we are not entitled to it under the rigorous exclusion of all notions but those of strict logic: Our values of M/V and the value of ρ must remain without link or bind.

Returning to collectives, let us consider how those who regard probability

[5] R. von Mises, l.c., p. 121.

[6] H. Reichenbach, *Wahrscheinlichkeitslehre*, Leiden 1935.

[7] It is understood, of course, that certain quantities of the sort here considered may be derived from others by the laws of the physical theory. We merely claim that there are certain initial points at which the present difficulty arises.

as a frequency have sought to deal with the present question.[8] The limit w is regarded as of the category of (idealized) physical quantities defined by limits such as ρ; the observational material $(\alpha_1 + \cdots + \alpha_N)/N$ is the analogue of M/V. And it is regarded that conclusions applying to $(\alpha_1 + \cdots + \alpha_N)/N$ can be drawn from w and *vice versa* in exactly the same sense and with the same validity as in the case of M/V and ρ: Why should a practical experimenter hesitate to do this more in the former case than in the latter?

This reasoning of course is impeccable. But it must be clear by now that it is two-edged: either it compels us to concede that some rudimentary notion of intuitive probability is presupposed in the applications of the collectivist theory, or else it merely assimilates one insurmountable difficulty to another.

Having reached the point where it is compulsory to regard the applied notion of frequency as presupposing the intuitive concept of probability, the question is raised as to precisely how the relation between the two is to be formulated. Perhaps the most obvious way of doing this and at the same time making the fewest demands on the intuition would be to enunciate the following principle: *The intuitive probability of success of E* (in (α)) *increases with w.*

Put in explicitly logical form and with the intuitive symbol $(\prec)$ this becomes

Hypothesis. (i) *The sequence (α) of trials of E is a collective of frequency w.* (ii) *The sequence (α^*) of trials of a second event E^* is a collective of frequency w^*.* (iii) $w < w^*$.

Conclusion. (E *succeeds on* n^{th} *trial*) $\prec$ (E^* *succeeds on* n^{th} *trial*).

This principle would enable us to derive non-trivial instances of $(\prec)$ from the hypothesis (i) (ii) (iii) which do not contain any non-trivial statements involving $(\prec)$ and thus place us in the presence of a situation utterly new in our theory, and quite contrary to our view expressed at the end of §1 of AAP to the effect that no such derivation is possible in the nature of things.

It is easy to show, however, that the suggested principle must be modified.

Case 1: suppose that in a particular instance of its application we are reliably informed of the outcome of some or all of the trials of E and E^*, i.e., that we know the values of certain α_n, α_n^*. Then evidently the conclusion would be absurd if we knew e.g. that $\alpha_n > \alpha_n^*$—whatever the values of w and w^*.

Case 2: suppose that we are informed of the results of certain of the outcomes by a person in whose reliability we place considerable but not absolute trust. If he has declared that $\alpha_n > \alpha_n^*$ this might lead us to deny the conclusion of the principle for the n'^{th} trial—and on perfectly reasonable grounds.

Case 3: suppose that we believe that certain more or less external events (weather, sun-spots, phases of the moon, incantations) are capable of influencing the outcome of E and E^*, being favorable let us say to E, unfavorable to E^*. Then we might well refrain from admitting the conclusion, particularly if the values of w and w^* are exceedingly close. A second person may regard our idea of an influence on E and E^* as reasonable, far-fetched, or absurd, depending on

[8] R. von Mises, l.c., p. 124.

the case in point, but he will never be able to prove that we are wrong by any reasoning or objective experiment which does not itself start by making *a priori* assumptions involving ($\prec$)—certainly not by the use of reasoning based on the suggested principle concerning sequences if a glaring *petitio* is to be avoided.

Evidently hypotheses (i) (ii) (iii) are insufficient to insure the conclusion: Something further must be added which is in the nature of an exclusion of the subsidiary knowledge (Case 1), reasonable certainty (Case 2), or disposition of belief (Case 3). Again, if we have made the conception of intuitive probability in the leastwise clear, the form which this additional hypothesis must take is evident: *It must make assertions in terms of* ($\prec$). The precise phrasing of the condition will incorporate the idea that the various trials of E are made under essentially the same circumstances, as well as the notion that the trials are either independent events or have a constant influence upon one another. This whole question is studied in §3, the final condition being what is there termed the *symmetry* of the sequence of trials. The final addition to the hypothesis thus becomes:

HYPOTHESIS. (iv) *If* $(i_1, \cdots, i_t)$ *and* $(j_1, \cdots, j_t)$ *are any two sets each containing t distinct positive integers, then the intuitive probability that the trials of E of orders* $i_1, \cdots, i_t$ *should all lead to success is equal* ($\approx$) *to that for trials* $j_1, \cdots, j_t$. *This is assumed for* $t = 1, 2, \cdots$.

(v) *Similarly for* E^*.

The question now arises whether under all these hypotheses (i) (ii) (iii) (iv) (v) the conclusion may reasonably be maintained, and if so, whether this fact constitutes a new principle of probability which requires to be posited. The answer is trenchant: Absolutely no new principle need be assumed, and the conclusion in question is the conclusion of a *theorem* which has as its hypothesis the assumptions (iii) (iv) (v) and the first half only of assumptions (i) and (ii). And that part of (i) and (ii) which is discarded is precisely Assumption 2 in the definition of collectives,—that very condition, namely, which has been found most unwelcome inasmuch as it involves the somewhat arbitrary choice of the class S of place-selection rules.

Before proceeding it is necessary to settle one question. Granting that $\lim (\alpha_1 + \cdots + \alpha_n)/n = w$, we may clearly rearrange the sequence (α) (at least when $0 < w < 1$) so that any other preassigned limit is obtained; does this not lead to inconsistent results? This question is answered in the negative by scrutinizing the hypothesis concerning the existence of the limit w in the light of deeper logical considerations, whereupon the hypothesis in question appears capable of two utterly distinct interpretations; the one presupposed in our theorem is not the one which permits this rearrangement. These logical preliminaries occupy us in §2.

One final remark regarding the precise formulation of this theorem in terms of the common presumption h: What right have we to assume in a given application that the concrete circumstances in the n'^{th} trial are the same as in the m'^{th}? It would clearly be more fitting to let the presumption in the first case be h_n,

that in the second be h_m. The answer to this question is that in the case in point we assume that a certain h can be formulated for which $a_n/h_n \approx a_n/h$ $(n = 1, 2, \cdots)$: this is part of the assumption that the trials are all made "under essentially similar conditions." While it is an assumption external to the theory of probability, being a statement of the observer's attitude towards a physical situation, its precise formulation is made possible in the language of the theory, which thus fulfills exactly its appropriate function and nothing more.

These remarks have bearing on Theorems 11 and 12 below.

2. A Logical Distinction

It has been explained (AAP §2) that the present theory makes use of two distinct classes of propositions. There are firstly the *contemplated* propositions, these being the concrete statements of the outcome of particular physical or biological events verifiable in principle by the performance of appropriate crucial experiments: They are denoted by lower case Latin letters and their finite $(\sim \cdot \vee)$ combinations. There are secondly the *asserted* propositions, which can always be denoted by connecting a pair of symbols for contemplated propositions by means of a single one of the assertive symbols $(\subset, \prec)$ or their derivates $(=, \approx, <, \|$, etc.). All the axioms and theorems of the present theory have a standard form: both hypothesis and conclusion are sets (finite or infinite) of asserted propositions; for in last analysis they are but statements of the laws of consistency governing any aggregate of logico-probability assertions which a given individual at a given moment can make.[9]

It will be convenient henceforth to depart from one convention of notation made hitherto: We will apply the logical symbols $(\sim \cdot \vee)$, as well as $\prod$ and $\sum$, to *asserted* propositions, thereby obtaining new asserted propositions *of the same logical type*. Thus for example

$$\text{(i)} \quad \sim (a/h \prec b/k); \qquad \text{(ii)} \quad (a = 1) \vee (b = 1);$$

$$\text{(iii)} \quad \prod_{n=1}^{\infty} (a_n = 1), \quad \text{i.e,} \quad (a_1 = 1)(a_2 = 1) \cdots ;$$

are regarded as mere abbreviations for assertions which may be couched respectively in the following intuitive-logical forms:

(i) The individual at the moment considered refuses to regard $a/h \prec b/k$.

(ii) Either it is asserted that a is true, or else it is asserted that b is true, or else that both a and b are true.

(iii) For each $n = 1, 2, \cdots$ it is asserted that a_n is true.

[9] Two comments may be made in passing: Firstly, in view of Axiom I, $(\subset)$ may be replaced throughout by $(\prec)$; for $a \subset b$ which coincides with $a \sim b \subset 0$ is coextensive with $a \sim b/1 \prec 0/1$. Secondly, in the assertion $a/h \prec b/k$, on the left h is in a sense asserted, on the right on the other hand, k is regarded as asserted. Such temporary or localized assertions are not of the kind which form hypotheses or conclusions in our axioms or theorems; that is why we have distinguished them by the appellation of *presumptions*.

It is to be emphasized that this notation is introduced purely for convenience as an abbreviation: without it all the forthcoming results would subsist. We shall assume the intuitive-logical rules for dealing with these symbols, and use parenthesis in a self-explanatory manner.[10]

We now come to the point where a certain logical distinction must be explicitly noted: *A disjunction of asserted propositions is fundamentally different from the assertion of the corresponding disjunction of the contemplated propositions.* Thus if a and b are contemplated propositions, the asserted proposition $(a = 1) \vee (b = 1)$ is fundamentally different from the asserted proposition $(a \vee b) = 1$. Similarly, the two following assertions are absolutely different.

$$\sum_{n=1}^{\infty} (a_n = 1); \qquad (\sum_{n=1}^{\infty} a_n) = 1.$$

Thus when $(a = 1) \vee (b = 1)$ occurs in the hypothesis of a theorem, we can always split the theorem up into two theorems, the first with $a = 1$ in the hypothesis, the second with $b = 1$: What is common to the conclusion of both theorems is then taken as the conclusion of the original theorem with the hypothesis $(a = 1) \vee (b = 1)$. Similarly, if $\sum_n(a_n = 1)$ is given by hypothesis, to prove the conclusion amounts to proving that it follows from each hypothesis $a_n = 1$ $(n = 1, 2, \cdots)$. This is essentially different if the hypothesis is $(a \vee b) = 1$ in the first case, $(\sum_n a_n) = 1$ in the second. This point is so fundamental in the foundations of probability and has been so generally ignored that we shall consider in detail some of the paradoxes to which its disregard may lead.

Firstly there is the frequently heard objection to the possibility of any theory of probability, typified by the remark that proposition a is either true or false; in either case such an assertion as $a/1 \approx \sim a/1$ is untrue; hence it is an impossible assertion. To be sure, we always assert that a is either true or false in the sense that we always assert $(a \vee \sim a) = 1$; but in the above reasoning this is fallaciously confused with $(a = 1) \vee (\sim a = 1)$ and since the conclusion $\sim(a/1 \approx \sim a/1)$ is a consequence of $a = 1$ as well as of $\sim a = 1$, its general validity is regarded as established. Instances of like kind lead us to the view that the distinction upon which we are dwelling is indeed a *sine qua non* in the theory of probability.

Secondly, one might attempt the following proof of the Axiom of Alternative presumption (AAP §3), or better, of the following generalization:

HYPOTHESIS. (i) $a/hb_n \prec r/s$ $(n = 1, 2, \cdots)$.

(ii) $h \subset \sum_{n=1}^{\infty} b_n$, i.e., $(h \sum_{n=1}^{\infty} b_n) = h$.

[10] Clearly we could carry the transcription into logical notation one step further and express all our axioms and theorems as implication statements between the $(\subset, \prec)$-assertions. Thus we would be in the presence of three distinct logical types of propositions: the contemplated propositions, the $(\subset, \prec)$-assertions, and the axioms and theorems of probability. But it is unnecessary and may be confusing to carry the symbolization so far in a work of this nature.

(iii) $b_i b_j = 0 (i \neq j)$.

CONCLUSION. $a/h \prec r/s$.

For (ii) can evidently be written

$$(\sum_{n=1}^{\infty} hb_n) = h,$$

which, if the fundamental distinction were not observed would be confused with

$$\text{(iii)} \quad \sum_{n=1}^{\infty} (hb_n = h), \text{ i.e., } \sum_{n=1}^{\infty} (h \subset b_n).$$

This latter can, as we have already explained, be split into the sequence of hypotheses $h \subset b_n$ $(n = 1, 2, \cdots)$. But in each case $hb_n = h$ and we have $a/h = a/hb_n \prec r/s$, and thus the conclusion, being true in each case, follows in general. The invalidity of this proof due to the confusion of (iii) with (ii) is what has forced us to list Axiom P as an unproved assumption. As a matter of fact the above infinite extension of Axiom P appears to be false, as the following example would show.

Consider two sequences of propositions $(a_1, a_2, \cdots)$ and $(a_1', a_2', \cdots)$ satisfying the conditions

$$\text{(1)} \qquad a = \sum_{n=1}^{\infty} a_n; \qquad a' = \sum_{n=1}^{\infty} a_n'; \qquad (a \vee a') = h \neq 0;$$

$$\text{(2)} \qquad \begin{aligned} a_i a_j &= a_i' a_j' = 0 && (i \neq j); \\ a_i a_j' &= 0 && (i, j = 1, 2, \cdots); \end{aligned}$$

$$\text{(3)} \qquad a_i/h \approx a_j/h \approx a_i'/h \approx a_j'/h.$$

No a_i or $a_j' = 0$ as this would lead by (3) and AAP Theorem 1 to this relation for all a_i, a_j', hence to $h = 0$ contrary to (1). Further, the evident relation $a_i \vee a_j' \vee a_k'/h \approx a_i \vee a_j' \vee a_k'/h$ in combination with (3) gives, by AAP Theorem 2 (i.e., Axiom D sharpened)

$$a_i/a_i \vee a_j' \vee a_k' \approx a_j'/a_i \vee a_j' \vee a_k' \approx a_k'/a_i \vee a_j' \vee a_k'$$

from which it appears (AAP Theorem 20) that $a_i/a_i \vee a_j' \vee a_k'$ is appraisable and of numerical probability 1/3. Hence, if r/s is appraisable and of numerical probability 5/12, we have $a_i/a_i \vee a_j' \vee a_k' \prec r/s$. On setting $b_n = a_n \vee a_{2n-1}' \vee a_{2n}'$ $(n = 1, 2, \cdots)$ and noting the relations $a/hb_n = ab_n/b_n = a_n/a_n a_{2n-1}' a_{2n}'$ we find that Hypothesis (i) of the theorem we are testing is verified; and (ii) is a consequence of (1), (iii) of (2). Hence the conclusion $a/h \prec r/s$, hence $p^*(a/h) \leqq 5/12$.

On replacing a etc. by a' etc. in the above reasoning, we obtain $p^*(a'/h) \leqq 5/12$. Hence by AAP Theorem 18 we have $p^*(h/h) = p^*(a \vee a'/h) \leqq p^*(a/h) + p^*(a'/h) \leqq 5/6$. And since $p^*(h/h) = 1$, we have a contradiction.

Returning to the sequence of trials of E, let a_n denote the contemplated experi-

mental proposition "the n'^{th} trial leads to success." We have now to cast into unambiguous logical form the definition of the statistical weight or frequency of success w. According to the First Assumption of Collectives: *If $r(n)$ is the number of the true propositions in the set $(a_1, \cdots, a_n)$, then* $\lim_{n\to\infty} r(n)/n = w$.

First of all, how is this definition of $r(n)$ to be understood? At the two extremes there are the following interpretations, denoted for brevity by $U(n, r)$ and $V(n, r)$:

$$U(n, r): \qquad \sum_{(s)} (a_{s_1} \cdots a_{s_r} \sim a_{s_{r+1}} \cdots \sim a_{s_n} = 1),$$

$$V(n, r): \qquad (\sum_{(s)} a_{s1} \cdots a_{s_r} \sim a_{s_{r+1}} \cdots \sim a_{s_n}) = 1.$$

Here $(s) = (s_1, \cdots, s_n)$ denotes any permutation of $(1, \cdots, n)$ and the $\sum_{(s)}$ calls for the formation of all the terms for all possible (s) and their combination by disjunction $(\vee)$: that only $n!/r!(n-r)!$ of the terms are distinct is immaterial. But between these two extremes are a host of intermediate possibilities; thus in the case $n = 3$, $r = 2$ there are formulations such as

$$[(a_1a_2 \sim a_3 \vee a_1a_3 \sim a_2) = 1] \vee [a_2a_3 \sim a_1 = 1].$$

It will be realized at once that $V(n, r)$ is in a certain sense the weakest of all these expressions; it turns out to be the natural one for the theory of probability. But there is still a further modification in the direction of weakening which can be made, as we shall now see.

The explicit wording of the definition of w would now appear to be: "For every positive integer μ there exists a positive integer m such that assertion $V(n, r)$ holds for all $n = m, m + 1, m + 2, \cdots$, and all r between $nw - n/\mu$ and $nw + n/\mu$." This suggests the precise logical rendering:

$$\prod_{\mu=1}^{\infty} \sum_{m=1}^{\infty} \prod_{n=m}^{\infty} \sum_{nw-n/\mu}^{nw+n/\mu} (r)\, V(n, r).$$

But this is capable of further weakening; for

$$\sum_{nw-n/\mu}^{nw+n/\mu} (r) \{(\sum_{(s)} a_{s_1} \cdots a_{s_r} \sim a_{s_{r+1}} \cdots \sim a_{s_n}) = 1\}$$

is only one possible interpretation of the statement: "the number of true propositions in the set $(a_1, \cdots, a_n)$ is between $nw - n/\mu$ and $nw + n/\mu$." Another rendering is the assertion which we will call $W(w, \mu, n)$ suggested by modifying the preceding formula into

$$\{\sum_{nw-n/\mu}^{nw+n/\mu} (r) \sum_{(s)} a_{s_1} \cdots a_{s_r} \sim a_{s_{r+1}} \cdots \sim a_{s_n}\} = 1.$$

Some evident algebraic manipulations allow it to be put into a simpler form. Let $t = t(w, \mu, n)$ denote the least integer $\geqq nw - n/\mu$, and $f = f(w, \mu, n)$ the least integer $\geqq n(1 - w) - n/\mu$. Let

$$(p, q) = (p_1, \cdots, p_t, q_1, \cdots, q_f)$$

denote any set of $t + f$ distinct integers chosen from $(1, \cdots, n)$. Finally, employ $\sum(p, q)$ to denote the disjunction $(\vee)$ of all the terms formed by taking all possible choices of (p, q) $(w, \mu, n$ being fixed) in the summand. Then the definitive formulation of the weakest interpretation of the assertion "the number of true propositions in $(a_1, \cdots, a_n)$ is between $nw - n/\mu$ and $nw + n/\mu$" is as follows

$$W(w, \mu, n): \ (\sum_{(p,q)} a_{p_1} \cdots a_{p_t} \sim a_{q_1} \cdots \sim a_{q_f}) = 1.$$

(Note that only $n!/t!f!(n - t - f)!$ terms in the disjunction are distinct). In terms of this we have the assertion characterizing w:

$$W(w): \prod_{\mu=1}^{\infty} \sum_{m=1}^{\infty} \prod_{n=m}^{\infty} W(w, \mu, n).$$

We are now in a position to consider the following paradox: When $0 < w < 1$ there are infinitely many true propositions in $(a_1, a_2, \cdots)$ and infinitely many false; let the former be denoted in their order of occurrence by $(a_{\tau_1}, a_{\tau_2}, \cdots)$ and the latter by $(a_{\varphi_1}, a_{\varphi_2}, \cdots)$. Then any number θ $(0 \leqq \theta \leqq 1)$ being given, an explicit process is easily formulated which combines these two sequences into a single sequence—which is but a re-ordering of $(a_1, a_2, \cdots)$ –for which the frequency of the a_{τ_i} is θ. How then can a frequency be taken as the numerical probability of success of E?

The whole question hinges on the meaning of a_{τ_1}. If the statement "a_{τ_1} is the first true proposition in $(a_1, a_2, \cdots)$" be taken to mean "true in Nature" i.e., as experimentally verifiable but of unknown verity, then the frequency of true propositions has nothing to do with the definition $W(w)$ of w. If it means "asserted as true," i.e., $= 1$, then $W(w)$ does not allow a_{τ_1} to be defined. For $W(w)$ allows us only to conclude that for some w, μ, n for which $t(w, \mu, n) > 0$, we have the assertion $W(w, \mu, n)$; in one sense this tells us that at least one a_i in the set $(a_1, \cdots, a_n)$ is true, but in precisely the same sense as the assertion $(a \vee \sim a) = 1$ tells us that at least one of $(a, \sim a)$ is true: it leaves the verbiage "the first proposition in $(a, \sim a)$ which is asserted as true" without meaning. In last analysis it is but the first paradox which we have considered in more complicated form.

If $W(w)$ were replaced by

$$\prod_{\mu=1}^{\infty} \sum_{m=1}^{\infty} \prod_{n=m}^{\infty} \sum_{(p,q)} (a_{p_1} \cdots a_{p_t} \sim a_{q_1} \cdots \sim a_{q_f} = 1)$$

the definition of $(a_{\tau_1}, a_{\tau_2}, \cdots)$ and $(a_{\varphi_1}, a_{\varphi_2}, \cdots)$ would be possible. Thus if $w = 1/2$, $\mu = 4$, and $n = 4\nu \geqq m$, we have

$$\sum_{(p,q)} (a_{p_1} \cdots a_{p_t} \sim a_{q_1} \cdots \sim a_{q_f} = 1).$$

Now t = least integer $\geqq nw - n/\mu = 2\nu - \nu = \nu$ so that $t = \nu$; similarly, $f = \nu$. Thus we have

$$(a_1 \cdots a_\nu \sim a_{\nu+1} \cdots \sim a_{2\nu} = 1) \vee \cdots \vee (a_{3\nu} \cdots a_{4\nu} \sim a_{2\nu} \cdots \sim a_{3\nu-1} = 1).$$

The disjoined terms are thought of as arranged lexicographically according to subscripts in the set of first factors such as $a_1 \cdots a_\nu$, which are themselves supposed to be written in increasing order of subscript. Thus we may speak of the first term asserted as true, i.e., such that $a_{p_1} \cdots a_{p_\nu} \sim a_{p_\nu+1} \cdots \sim a_{p_{2\nu}} = 1$, and then set $\tau_1 = p_1 : a_{\tau_1}$ is defined. Similarly for later terms.

Thus this paradox like the rest is resolved by maintaining the logical distinction which is the subject of this section.

Before closing we shall consider an extension. Let h be the common presumption made in envisaging the outcomes of E: We are considering the sequence of eventualities a_n/h. We wish to formulate an assertion $W_h(w)$ which is the extension of $W(w)$ and characterizes w regarded as the frequency of successes of E on the presumption that h is true. Such an assertion is evidently furnished by replacing $(\cdots) = 1$ in the above by the assertion $h \subset (\cdots)$ (or equivalently $h\cdot(\cdots) = h$). Everything else being as above, we set

$$W_h(w, \mu, n)\colon\ h \subset \sum_{(p,q)} a_{p_1} \cdots a_{p_t} \sim a_{q_1} \cdots \sim a_{q_f}$$

and correspondingly,

$$W_h(w)\colon\ \prod_{\mu=1}^{\infty} \sum_{m=1}^{\infty} \prod_{n=m}^{\infty} W_h(w, \mu, n).$$

Finally, if $\sum_{0 \leqq w \leqq 1}$ is used for the disjunction of all the terms formed by letting w range from 0 to 1, the assertion: "there is a frequency of successes on the presumption h" shall be formulated as

$$W_h\colon\ \sum_{0 \leqq w \leqq 1} W_h(w).$$

3. Symmetry and Independence

DEFINITION 1. *Any finite or countable set of propositions $(a_1, a_2, \cdots)$ shall be said to be a symmetric set with respect to the proposition $h(\neq 0)$ when for every possible positive integral t*

$$a_{i_1} \cdots a_{i_t}/h \approx a_{j_1} \cdots a_{j_t}/h \tag{1}$$

holds for every choice of the t distinct integers $(i_1, \cdots, i_t)$ and the t distinct integers $(j_1, \cdots, j_t)$.

Here $h \neq 0$ denotes the assertion $\sim (h = 0)$.

Let $\varphi(x_1, \cdots, x_m)$ denote a finite combination of the symbols $x_1, \cdots, x_m$ with the aid of the logical constants $(\sim \cdot \vee)$, the convention being that the formal laws of Boolean algebra hold for these symbols; such a $\varphi(x_1, \cdots, x_m)$ may be called a Boolean polynomial in these m (propositional) variables. If by

application of the formal Boolean identities alone, i.e., without assuming any individual properties of $x_1, \cdots, x_m$, it may be reduced to 0, 1, or some other given proposition not involving the variables, we will say that it is identically equal ($\equiv$) to this proposition.

THEOREM 1. *If $(a_1, a_2, \cdots)$ is a symmetric set with respect to h, then for any $\varphi(x_1, \cdots, x_m)$ (m not exceeding the number of propositions in the set) the relation*

$$\varphi(a_{i_1}, \cdots, a_{i_m})/h \approx \varphi(a_{j_1}, \cdots, a_{j_m})/h \tag{2}$$

holds for every possible pair $(i_1, \cdots, i_m)$ and $(j_1, \cdots, j_m)$ of m distinct positive integers each.

Elementary algebraic manipulations show that any $\varphi(x_1, \cdots, x_m)$ can be expressed as the disjunction of disjoint monomials $x_{k_1} \cdots x_{k_t} \sim x_{k_{t+1}} \cdots \sim x_{k_{t+s}}$; thus when (2) has been established for monomial $\varphi(x_1, \cdots, x_m)$ its general validity becomes an immediate consequence of the theorem of total probability (AAP Theorem 6, sharpened). The relation

$$a_{i_1} \cdots a_{i_t} \sim a_{i_{t+1}} \cdots \sim a_{i_{t+s}}/h \approx a_{j_1} \cdots a_{j_t} \sim a_{j_{t+1}} \cdots \sim a_{j_{t+s}}/h$$

is proved by induction for all s. Assume it true for all values of the first subscript t and all values of the second between 0 and s inclusive and apply AAP Theorem 7 (sharpened) to

$$a_{i_1} \cdots a_{i_t} \sim a_{i_{t+1}} \cdots \sim a_{i_{t+s}} a_{i_{t+s+1}}/h \approx a_{j_1} \cdots a_{j_t} \sim a_{j_{t+1}} \cdots \sim a_{j_{t+s}} a_{j_{t+s+1}}/h,$$

$$(a_{i_1} \cdots a_{i_t} \sim a_{i_{t+1}} \cdots \sim a_{i_{t+s}}, \qquad a_{i_1} \cdots a_{i_t} \sim a_{i_{t+1}} \cdots \sim a_{i_{t+s}} a_{i_{t+s+1}}, h$$

respectively replacing a_1, b_1, h_1, etc.). The required relation is an immediate consequence.

COROLLARY. *If $(a_1, a_2, \cdots)$ is symmetric with respect to h, then any subset of $(a_1, a_2, \cdots)$ and any subset of $(\sim a_1, \sim a_2, \cdots)$ is likewise.*

DEFINITION 2. *Any finite or countable set of propositions $(a_1, a_2, \cdots)$ shall be said to be an independent set with respect to the proposition $h(\neq 0)$ if firstly $ha_i \neq 0$ or h, and secondly*

$$a_i/ha_{i_1} \cdots a_{i_t} \approx a_i/h \tag{3}$$

for every possible t and every choice of $t + 1$ distinct integers $(i, i_1, \cdots, i_t)$.

We must prove that (3) has meaning in every case, i.e., that $ha_{i_1} \cdots a_{i_t} \neq 0$. This is assumed in the hypothesis when $t = 1$. Proceeding inductively, let it be assumed for the value t; we then will show that $ha_{i_1} \cdots a_{i_{t+1}} \neq 0$. If $ha_{i_1} \cdots a_{i_{t+1}} = 0$ the following relations would occur:

$$a_{i_{t+1}}/h \approx a_{i_{t+1}}/ha_{i_1} \cdots a_{i_t} = ha_{i_1} \cdots a_{i_{t+1}}/ha_{i_1} \cdots a_{i_t} = 0/ha_{i_1} \cdots a_{i_t}$$

and hence $ha_{i_{t+1}} = 0$ (AAP Theorem 1) contrary to hypothesis.

A precisely similar process shows that $ha_{i_1} \cdots a_{i_t} \neq h$, and furthermore that if $(i_1, \cdots, i_{t+s})$ are any $t + s$ distinct positive integers

$$ha_{i_1} \cdots a_{i_t} \sim a_{i_{t+1}} \cdots \sim a_{i_{t+s}} \neq 0 \text{ or } h.$$

From this relation Theorem 2 follows immediately:

THEOREM 2. *If* $(a_1, a_2, \cdots)$ *is an independent set with respect to* h *and if* $\varphi(x_1, \cdots, x_m)$ *is any Boolean polynomial not identically* 0 *or* h, *then for every* m *distinct* $(i_1, \cdots, i_m)$, $h\varphi(a_{i_1}, \cdots, a_{i_m}) \neq 0$ *or* h.

THEOREM 3. *If* $(a_1, a_2, \cdots)$ *is an independent set with respect to* h *and if* $\varphi(x_1, \cdots, x_m)$ $(\neq 0, h)$ *is as above, then for any* $m + 1$ *distinct* $(i, i_1, \cdots, i_m)$

$$a_i/h\varphi(a_{i_1}, \cdots, a_{i_m}) \approx a_i/h. \tag{4}$$

As in the proof of Theorem 1, $\varphi(x_1, \cdots, x_m)$ is expressed as the disjunction of disjoint monomials $x_{k_1} \cdots x_{k_t} \sim x_{k_{t+1}} \cdots \sim x_{k_{t+s}}$. Once the relation

$$a/ha_{i_1} \cdots a_{i_t} \sim a_{i_{t+1}} \cdots \sim a_{i_{t+s}} \approx a_i/h \tag{5}$$

has been established the truth of (4) will follow by an immediate inductive application of Axiom P.[11]

We assume the truth of (5) for all values of the first index t and all values of the second between 0 and s inclusive, and seek to prove it when the second index is replaced by $s + 1$. On account of the assumption we have

$$a_i/ha_{i_1} \cdots a_{i_t} \sim a_{i_{t+1}} \cdots \sim a_{i_{t+s}} a_{i_{t+s+1}} \approx a_i/h$$
$$\approx a_i/ha_{i_1} \cdots a_{i_t} \sim a_{i_{t+1}} \cdots \sim a_{i_{t+s}}.$$

The desired relation follows on applying AAP Theorem 10 (sharpened) with a_i, $a_{i_{t+s+1}}$, $ha_{i_1} \cdots a_{i_t} \sim a_{i_{t+1}} \cdots a_{i_{t+s}}$ replacing the a, k, h, respectively, of that theorem.

COROLLARY. *If* $(a_1, a_2, \cdots)$ *is an independent set with respect to* h, *then any subset of* $(a_1, a_2, \cdots)$ *and any subset of* $(\sim a_1, \sim a_2, \cdots)$ *is likewise.*

The notion of symmetry corresponds with that of events all equally probable and such that any combination of results of one set is as probable of occurrence as the same combination of another set. The notion of independence on the other hand corresponds with the idea that knowledge of the occurrence of some of the events does not change the probability ascribed to that of the others. A set can be symmetric without being independent, for symmetry does not exclude $h\varphi(a_1, \cdots, a_m) = 0, h$: sets in which $a_1 = a_2 = \cdots$ or in which $a_1 \vee \cdots \vee a_m = 0$ may be symmetric but can not be independent. On the other hand sets may be independent without being symmetric; for example a_1, a_2, etc. may denote success of respective trials of unconnected events but events of different kind and thus in general, different probability. Hence the significance of the following

---

[11] The theorem employed here is the following:

If $hc_ic_j = 0$ *for all* $i \neq j$ *and if* $a/hc_i \prec r/s$ $(i = 1, \cdots, n)$, *then* $a/hc \prec r/s$ *where* $c = c_1 \vee \cdots \vee c_n$.

This is proved inductively from the case $n = 2$, in which it is a self-evident consequence of Axiom $P(c_1$ and $h(c_1 \vee c_2)$ replacing b and $h)$. The sharpening and strengthenings are immediate. This theorem belongs properly in AAP §4, just before Theorem 13.

THEOREM 4. *If* $(a_1, a_2, \cdots)$ *is an independent set with respect to h and if* $a_i/h \approx a_j/h$ *for all i and j, then* $(a_1, a_2, \cdots)$ *is a symmetric set with respect to h.*

For we have by hypothesis

$$a_{i_1}/ha_{i_2} \approx a_{i_1}/h \approx a_{i_2}/h \approx a_{j_2}/h \approx a_{j_1}/h \approx a_{j_1}/haj_2 .$$

On applying Axiom C_1 (sharpened by AAP Theorem 2) with h, a_{i_2}, a_{i_1}, h, a_{j_2}, a_{j_1} in lieu of h_1, a_1, b_1, h_2, a_2, b_2, the result $a_{i_1}a_{i_2}/h \approx a_{j_1}a_{j_2}/h$ is obtained; the proof of the general relation (1) proceeds by induction.

In the classical theory the only relations of probability are those concerned with numerical probability. The notions of symmetry and independence are *less* restictive insofar as equiprobability $(a/h \approx b/k)$ implies equal numerical probability $(p(a/h) = p(b/k))$ and not conversely; but they are *more* restrictive in that the appraisability of every eventuality is assumed. We give the following definitions and theorems (in essence familiar) mainly for completeness. The proofs are supplied immediately by means of the laws of total and compound probability (cf. AAP, §5, Theorems 23 to 28).

Let $(a_1, a_2, \cdots)$ be any finite or countable set of propositions, let $h \neq 0$ be a further proposition, and assume that all the eventualities with respect to h (i.e., remainder classes with respect to the principle ideal $(\sim h)$ in the Boolean ring determined by $(h_1, a_1, a_2, \cdots))$ are appraisable.

DEFINITION 3. *The set* $(a_1, a_2, \cdots)$ *shall be called numerically symmetric with respect to h when for all indices as in Definition* 1,

$$p(a_{i_1} \cdots a_{i_t}/h) = p(a_{j_1} \cdots a_{j_t}/h). \tag{5}$$

THEOREM 5. *If* $(a_1, a_2, \cdots)$ *is a numerically symmetric set with respect to h, then for any Boolean polynomial* $\varphi(x_1, \cdots, x_m)$

$$p(\varphi(a_{i_1}, \cdots, a_{i_m})/h) = p(\varphi(a_{j_1}, \cdots, a_{j_m})/h).$$

COROLLARY. *The numerical symmetry with respect to h of* $(a_1, a_2, \cdots)$ *implies that of any sub-set of* $(a_1, a_2, \cdots)$ *and any sub-set of* $(\sim a_1, \sim a_2, \cdots)$.

DEFINITION 4. *The set* $(a_1, a_2, \cdots)$ *shall be said to be numerically independent with respect to h if firstly* $0 < p(a_i/h) < 1$ *for all i, and secondly (the indices being as in Definition* 2),

$$p(a_i/ha_{i_1} \cdots a_{i_t}) = p(a_i/h). \tag{6}$$

THEOREM 6. *If* $(a_1, a_2, \cdots)$ *is numerically independent with respect to h and* $\varphi(x_1, \cdots, x_m)$ *any Boolean polynomial* ($\not\equiv 0, h$) *then for every m distinct* $(i_1, \cdots, i_m)$

$$0 < p(\varphi(a_{i_1}, \cdots, a_{i_m})/h) < 1.$$

THEOREM 7. *If* $(a_1, a_2, \cdots)$ *and* $\varphi(x_1, \cdots, x_m)$ *are as in Theorem* 6, *then for any* $m + 1$ *distinct* $(i, i_1, \cdots, i_m)$

$$p(a_i/h\varphi(a_{i_1}, \cdots, a_{i_m})) = p(a_i/h).$$

COROLLARY 1. *Under these conditions* $p(a_{i_1} \cdots a_{i_m}/h) = p(a_{i_1}/h)p(a_{i_2}/h) \cdots p(a_{i_m}/h)$.

COROLLARY 2. *The numerical independence with respect to h of* $(a_1, a_2, \cdots)$ *implies that of any sub-set of* $(a_1, a_2, \cdots)$ *and of any sub-set of* $(\sim a_1, \sim a_2, \cdots)$.

COROLLARY 3. *If* $(a_1, a_2, \cdots)$ *is a numerically independent set with respect to h and if for all i and j,* $p(a_i/h) = p(a_j/h)$, *then it is numerically symmetric.*

4. Theorems on Sequences

THEOREM 8. *If* $(a_1, \cdots, a_n)$ *is symmetric with respect to h and if precisely* $r(0 \leqq r \leqq n)$ *of its elements are implied by h in the sense that* (§2)

$$V_h(n, r):\ h \subset \sum_{(s)} a_{s_1} \cdots a_{s_r} \sim a_{s_{r+1}} \cdots \sim a_{s_n},$$

then a_i/h *is appraisable and* $p(a_i/h) = r/n$ $(i = 1, \cdots, n)$.

There is no loss of generality in assuming that $a_i \subset h$, i.e., $ha_i = a_i$ $(i = 1, \cdots, n)$; for this property can always be secured by replacing $(a_1, \cdots, a_n)$ by $(ha_1, \cdots, ha_n)$ which evidently also satisfies the hypothesis; since $ha_i/h = a_i/h$ the conclusions are the same. Proceeding on this assumption and excluding the trivial case of our theorem in which $r = 0$, we have on multiplying hypothesis $V_h(n, r)$ through by h

$$h = \sum_{(s)} a_{s_1} \cdots a_{s_r} \sim a_{s_{r+1}} \cdots \sim a_{s_n}.$$

The summation $\sum_{(s)}$ is for all permutations $(s) : (s_1, \cdots, s_n)$ of $(1, \cdots, n)$. as above.

There are evidently $\nu = n!/r!(n - r)!$ distinct terms in this equation (i.e., terms not reducible to one another by the identities of Boolean algebra); let them be denoted in some conventional order by $(u_1, \cdots, u_\nu)$. Of these terms there will be clearly $\tau = (n - 1)!/(r - 1)!\,(n - r)!$ distinct ones involving a preassigned a_i as factor; let them be $(u_{i_1}, \cdots, u_{i_\tau})$. We have moreover $u_k u_l = 0$ for all $k \neq l$, and of course $u_1 \vee \cdots \vee u_\nu = h \neq 0$. Finally, as a result of the symmetry (cf. Theorem 1) $u_k/h \approx u_l/h$ for all k and l. Thus $(u_1, \cdots, u_\nu)$ forms a ν-scale (AAP, §5, Definition 1) and if we set $b = u_{i_1} \vee \cdots \vee u_{i_\tau}$ we have by AAP Theorem 20 that b/h is appraisable and $p(b/h) = \tau/\nu = r/n$. Thus our theorem will be proved once it is established that $b/h = a_i/h$, i.e., that $hb = ha_i$. This is shown on multiplying the equation

$$u_1 \vee \cdots \vee u_\nu = b \vee u_{j_1} \vee \cdots \vee u_{j_{\nu-\tau}} = h$$

through by a_i : clearly if u_{j_1} does not have a_i as a factor, it does have $\sim a_i$, so that $a_i u_{j_1} = 0$, etc.

COROLLARY. *If* $(a_1, \cdots, a_n)$, *h and* $(a_1', \cdots, a_n')$, *h' each satisfy the hypothesis of Theorem 8 for the same value of r, then* $a_i/h \approx a_i'/h'$.

THEOREM 9. *If* $(a_1, \cdots, a_n)$ *is symmetric with respect to h and if at least r* $(0 \leqq r \leqq 1)$ *of its elements are implied by h in the sense that*

$$X_h(n, r):\ h \subset \sum_{(s_1, \cdots, s_r)} a_{s_1} \cdots a_{s_r}$$

(where the summation takes in all possible sets of r distinct integers in $(1, \cdots, n)$*), then for each value of i,* $p_*(a_i/h) \geqq r/n$.

Again we assume $a_i \subset h$ $(i = 1, \cdots, n)$ and disregard the trivial case $r = 0$. Multiplication by h puts $X_h(n, r)$ into the form $h = \sum a_{s_1} \cdots a_{s_r}$. We set

$$c_\rho = \sum_{(s)} a_{s_1} \cdots a_{s\rho} \sim a_{s\rho+1} \cdots \sim a_{s_n}; \tag{1}$$

here as in $V_h(n, r)$ the summation is for all permutations (s): $(s_1, \cdots, s_n)$ of $(1, \cdots, n)$. Simple algebraic manipulations establish

$$h = \sum_{\rho=r}^{n} c_\rho; \qquad c_\rho c_\sigma = 0 \ (\rho \neq \sigma); \qquad hc_\rho = c_\rho. \tag{2}$$

Let ρ be a value for which $c_\rho \neq 0$: since $h \neq 0$, such values will occur. We will show that $(a_1, \cdots, a_n)$ is symmetric with respect to c_ρ. In order to prove

$$a_{i_1} \cdots a_{i_t}/c_\rho \approx a_{j_1} \cdots a_{j_t}/c_\rho \tag{3}$$

we first introduce the Boolean polynomial

$$\varphi(x_1, \cdots, x_n) \equiv x_1 \cdots x_t \sum_{(s)} x_{s_1} \cdots x_{s_\rho} \sim x_{s_{\rho+1}} \cdots \sim x_{s_n}$$

and apply Theorem 1, whereupon the equation

$$\varphi(x_{i_1}, \cdots, x_{i_t}, x_{i_{t+1}}, \cdots, x_{i_n})/h \approx \varphi(x_{j_1}, \cdots, x_{j_t}, x_{j_{t+1}}, \cdots, x_{j_n})/h$$

is obtained (all subscripts in each set being distinct). But this may be written as

$$c_\rho a_{i_1} \cdots a_{i_t}/h \approx c_\rho a_{j_1} \cdots a_{j_t}/h. \tag{4}$$

Now if either $c_\rho a_{i_1} \cdots a_{i_t} = 0$ or $c_\rho a_{j_1} \cdots a_{j_t} = 0$, (4) shows that both these equations hold, and (3) becomes evident. In all other cases we apply Axiom D (sharpened) to (4) together with $c_\rho/h \approx c_\rho/h$ (with $a_1 = a_2 = c_\rho$, $h_1 = h_2 = h$, $b_1 = a_{i_1} \cdots a_{i_t}$, $b_2 = a_{j_1} \cdots a_{j_t}$) whereupon (3) is established.

This symmetry and the relation (1) show that $(a_1, \cdots, a_n)$ satisfies the hypothesis of Theorem 8 with (ρ, c_ρ) replacing (r, h); it follows that a_i/c_ρ is appraisable and $p(a_i/c_\rho) = \rho/n$ $(i = 1, \cdots, n)$.

Let σ be any given positive integer, and $(u_1, \cdots, u_{n\sigma})$ an $n\sigma$-scale. Set $v = u_1 \vee \cdots \vee u_{r\sigma-1}$, $u = u_1 \vee \cdots \vee u_{n\sigma}$; then $p(v/u) = (r\sigma - 1)/n\sigma = r/n - 1/n\sigma$ (AAP Theorem 20). Hence for each ρ for which $c_\rho \neq 0$ we have $a_i/c_\rho \succ v/u$ (AAP Theorem 16).

Returning to (2), we may write $h = \sum' c_\rho$ where the summation $\sum'$ includes only those terms of $(c_r, \cdots, c_n)$ which $\neq 0$. By applying Axiom P[11] to $a_i/c_\rho \succ v/u$ we derive the result that $a_i/\sum' c_\rho = a_i/h \succ v/u$. Hence by definition of $p_*(a_i/h)$ we have for all values of the positive integer σ that $p_*(a_i/h) \geqq r/n - 1/n\sigma$, which leads at once to the conclusion of our theorem.

THEOREM 10. *If* $(a_1, \cdots, a_n)$ *is symmetric with respect to h and if h implies the truth of at least t and falsehood of at least f of its elements in the sense that*

$$X_h(n, t, f)\colon h \subset \sum_{(p,q)} a_{p_1} \cdots a_{p_t} \sim a_{q_1} \cdots \sim a_{q_f}$$

(where the summation takes in all possible sets of $t + f$ distinct integers from $(1, \cdots, n)$), then

$$\frac{t}{n} \leqq p_*(a_i/h) \leqq p^*(a_i/h) \leqq 1 - \frac{f}{n}. \tag{5}$$

The first inequality in (5) results from the application of Theorem 9 ($r = t$) inasmuch as it is clear that

$$h \subset \sum_{(p,q)} a_{p_1} \cdots a_{p_t} \sim a_{q_1} \cdots \sim a_{q_f} \subset \sum_{(s)} a_{s_1} \cdots a_{s_t}.$$

To obtain the last inequality in (5) we have but to replace $(a_1, \cdots, a_n)$ by $(\sim a_1, \cdots, \sim a_n)$ (cf. Theorem 1, Corollary) and apply AAP Theorem 17 to the resulting inequality

$$f/n \leqq p_*(\sim a_i/h) = 1 - p^*(a_i/h).$$

THEOREM 11. *Let the infinite sequence $(a_1, a_2, \cdots)$ be a symmetric set with respect to h, and let the limiting frequency (statistical weight) of its elements implied by h be w in the sense of the assertion*

$$W_h(w)\colon \prod_{\mu=1}^{\infty} \sum_{m=1}^{\infty} \prod_{n=m}^{\infty} W_h(w, \mu, n),$$

in which $W_h(w, \mu, n)$ coincides with the assertion $X_h(n, t, f)$ in Theorem 10 in which $t = t(w, \mu, n)$ is the least integer $\geqq nw - n/\mu$ and $f = f(w, \mu, n)$ is the least integer $\geqq n(1 - w) - n/\mu$.

Then a_i/h is appraisable and $p(a_i/h) = w$ $(i = 1, 2, \cdots)$.

Let us choose a value of μ and hold it fast. Then in virtue of the hypothesis $W_\mu(w)$ there exists an integer m for which $\prod_{n=m}^{\infty} W_h(w, \mu, n)$ is valid. For definiteness we select the least m for which this is true (we may as well represent it by m); then we have for this m and, e.g., for $n = m$ the assertion $W_h(w, \mu, n)$. Hence by Theorem 10 we conclude that

$$\frac{t(w, \mu, n)}{n} \leqq p_*(a_i/h) \leqq p^*(a_i/h) \leqq 1 - \frac{f(w, \mu, n)}{n}.$$

This yields on referring to the definitions of $t(w, \mu, n)$ and $f(w, \mu, n)$

$$w - \frac{1}{\mu} \leqq p_*(a_i/h) \leqq p^*(a_i/h) \leqq w + \frac{1}{\mu}.$$

This equation, being true for every $\mu = 1, 2, \cdots$, leads at once to the conclusion of our theorem.

COROLLARY. *Let $(a_1, a_2, \cdots)$ be symmetric with respect to h and let the limiting frequency of its elements implied by h exist and lie between w_1 and w_2 in the sense of the assertion $\sum_{w_1 \leqq w \leqq w_2} W_h(w)$. Then a_i/h is appraisable and $w_1 \leqq p(a_i/h) \leqq w_2$ $(i = 1, 2, \cdots)$.*

THEOREM 12. *Let* $(a_1, a_2, \cdots)$ *be symmetric with respect to h and* $(b_1, b_2, \cdots)$ *be symmetric with respect to k. Let the number of a's implied by h eventually become and remain greater than the number of b's implied by k in the sense of the assertion*

$$\Omega: \qquad \sum_{m=0}^{\infty} \prod_{n=m}^{\infty} \sum_{r=0}^{n-1} \{(h \subset \sum a_{s_1} \cdots a_{s_{r+1}})(k \subset \sum \sim b_{t_1} \cdots \sim b_{t_{n-r}})\}.$$

Then for every i, j: $a_i/h \succ b_j/k$.

As usual $(s_1, \cdots, s_{r+1})$ is a set of $r + 1$ distinct integers chosen from $(1, \cdots, n)$, all possible sets of this kind being taken in the $\sum$; similarly for $(t_1, \cdots, t_{n-r})$ and the corresponding $\sum$.

Let m denote an integer for which $\prod_{n=m}^{\infty} \cdots$ is true by assertion (for definiteness, the *least m*). Let n denote the greatest of (m, i, j). With this determination of n we have the assertion $\sum_{r=0}^{n-1} \{\ \}$. Let r denote a value for which assertion $\{\ \}$ is made. Then we have simultaneously

$$h \subset \sum a_{s_1} \cdots a_{s_{r+1}}, \qquad k \subset \sum \sim b_{t_1} \cdots \sim b_{t_{n-r}}.$$

We are now in a position to apply Theorem 9 to each of the sets $(a_1, \cdots, a_n)$ and $(\sim b_1, \cdots, \sim b_n)$, the latter of which is symmetric by the corollary to Theorem 1, and so obtain

$$p_*(a_i/h) \geqq \frac{r+1}{n}, \qquad p_*(\sim b_j/k) \geqq \frac{n-r}{n}.$$

But $p_*(\sim b_j/k) = 1 - p^*(b_j/k)$, and hence $p_*(a_i/h) > p^*(b_j/k)$, from which the conclusion of our theorem follows (AAP Theorems 17 and 16).

In concluding this section we may consider the question of replacing in the hypotheses of these theorems the requirement of symmetry by that of numerical symmetry. At first sight such a paraphrase would have the advantage of weakening the condition of symmetry; but this advantage is in our opinion far offset by the disadvantage that appraisability has to be assumed in the hypothesis—and it is indeed one of the chief functions of the theorems of this section that they afford the property of appraisability in their conclusions. Furthermore, symmetry rather than numerical symmetry appears to be the natural assumption to make in the applications. Nevertheless the quantitative conclusions subsist when numerical symmetry is assumed instead of symmetry. We confine ourselves to a statement of the following examples. The proofs, which proceed along rather similar lines to those of the earlier theorems and with the use of the classical properties of numerical probability, are omitted.

THEOREM 13. *If* $(a_1, \cdots, a_n)$ *is numerically symmetric with respect to h and if assertion* $X_h(n, t, f)$ *is made, then* $t/n \leqq p(a_i/h) \leqq 1 - f/n$.

THEOREM 14. *If* $(a_1, a_2, \cdots)$ *is numerically symmetric with respect to h and if assertion* $W_h(w)$ *is made, then* $p(a_i/h) = w$.

THEOREM 15. *If* $(a_1, a_2, \cdots)$ *is numerically symmetric with respect to h and*

$(b_1, b_2, \cdots)$ *is numerically symmetric with respect to* k *and if assertion* Ω *is made, then* $p(a_i/h) \geqq p(b_j/k)$.

In the theorems of this section the symmetry rather than the independence of the sets of propositions $(a_1, a_2, \cdots)$ appears to be the important notion. That such an appearance is deceptive when regard is had to the applications (the trials of event E) is realized when the possibility is considered that the symmetry of $(a_1, a_2, \cdots)$ with respect to h does not guarantee that of $(a_{n+1}, a_{n+2}, \cdots)$ with respect to $ha_1 \cdots a_n$: i.e., the results of the first n trials of E being known, we may not be in a position to apply our theorems to the others. Thus it would appear that the hypothesis of Theorem 4 affording both symmetry and independence is prescribed in such applications.

COLUMBIA UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

OPERATOR-THEORETICAL TREATMENT OF MARKOFF'S PROCESS AND MEAN ERGODIC THEOREM

BY KÔSAKU YOSIDA AND SHIZUO KAKUTANI

(Received November 15, 1939)

CHAPTER I. INTRODUCTION

It is the purpose of the present paper to give a consistent treatment of the problems of Markoff's process and mean ergodic theorem from the standpoint of the theory of iteration of bounded linear operations in Banach spaces.

Let Ω be an abstract space where a measure of Lebesgue type is defined, and consider a simple Markoff's process, by which a moving point $t \in \Omega$ is transferred stochastically inside Ω. If we denote by $P(t, E)$ the transition probability that a moving point $t \in \Omega$ is transferred into a Borel set E of Ω after the elapse of a unit-time, then we have always $P(t, E) \geqq 0$ and $P(t, \Omega) = 1$. We shall further assume that $P(t, E)$ is completely additive as a set function of Borel sets E if t is fixed, and that $P(t, E)$ is Borel measurable in t if E is fixed. Under these assumptions, the probability $P^{(n)}(t, E)$ that a moving point $t \in \Omega$ is transferred into a Borel set E of Ω after the elapse of n unit-times is given recurrently by

$$P^{(n)}(t, E) = \int_\Omega P^{(n-1)}(t, ds)P(s, E), \qquad n = 2, 3, \cdots; \qquad P^{(1)}(t, E) = P(t, E),$$

where the integration is of Radon-Stieltjes type.

The problem of Markoff consists in the investigation of the behavior of the sequence of the iterations $P^{(n)}(t, E)$ and their arithmetic means $Q^{(n)}(t, E) = \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(t, E)$ for large n.

In the special case, when Ω is composed of only a finite number ($= N$) of points $\left(\text{with equal measure } \frac{1}{N}\right)$, our Markoff's process is reduced to the classical one, and our problem is nothing but the investigation of the behavior of the iterations P^n of a matrix $P = (p_{ij})$, $i, j = 1, 2, \cdots, N$, such that $p_{ij} \geqq 0$, $i, j = 1, 2, \cdots, N$, and $\sum_{j=1}^{N} p_{ij} = 1$, $i = 1, 2, \cdots, N$. This case was first treated by A. Markoff in his classical papers, and in the last ten years many contributions to this problem were given by J. Hadamard, M. Fréchet, R. von Mises, A. Kolmogoroff, W. Doeblin, J. L. Doob and others. These results are all collected in a recent monograph of M. Fréchet [3], which is so complete that we find almost no need of giving further discussions on this case. We shall only remark that this case may be treated exactly in the same manner as in the general case.

In the general case, when Ω is a continuum,[1] this problem was discussed by

[1] We shall confine ourselves in this paper only to the case when Ω is the closed interval $0 \leqq t \leqq 1$. This is not an essential restriction.

B. Hostinsky, M. Fréchet, J. L. Doob, W. Doeblin and N. Kryloff-N. Bogolioùboff. B. Hostinsky [1] treated the case when $P(t, E)$ is given by a continuous density $p(t, s)$: $P(t, E) = \int_E p(t, s)\, ds$, where $p(t, s)$ is a continuous function of two variables t and s in $\Omega \times \Omega$; M. Fréchet [1], [2] considered the case when $p(t, s)$ is bounded and measurable instead of being continuous; J. L. Doob [1] weakened this condition by requiring only that $p(t, s)$ is uniformly (in t) integrable in s, that is, that there exists for any $\epsilon > 0$ an $\eta > 0$ such that $\text{mes}(E) < \eta$ implies $\int_E p(t, s)\, ds < \epsilon$ for any $t \in \Omega$; and lastly, W. Doeblin [1] discussed the case when the existence of the density $p(t, s)$ is not necessarily assumed. He has only assumed that

$$\text{(D)} \qquad \begin{cases} \textit{there exist an integer } d \textit{ and positive constants } b,\ \eta \textit{ such that } \text{mes}(E) < \eta \\ \textit{implies } P^{(d)}(t, E) \leqq 1 - b \textit{ for any } t \in \Omega. \end{cases}$$

It is clear that the condition (D) of W. Doeblin is more general than that of J. L. Doob, and under this condition, W. Doeblin has obtained, by a direct (point-set-theoretical) method, a considerably precise result.

In the present paper, we shall discuss these problems by another method, which is due to N. Kryloff-N. Bogolioùboff [1], [2], and which may be called an operator-theoretical method. Indeed, we shall treat these problems by considering $P^{(n)}(t, E)$ as a kernel of an integral operator in some special Banach spaces. This may be done in two ways: firstly,

$$x \to T(x) = y: \quad y(E) = \int_\Omega x(dt) P(t, E)$$

is a bounded positive[2] linear operation which maps the Banach space $(\mathbf{M})$[3] into itself; and secondly,

$$x \to \boldsymbol{T}(x) = y: \quad y(t) = \int_\Omega P(t, ds) y(s)$$

is a bounded positive linear operation which maps the Banach space (M^*)[4]

---

[2] A bounded linear operation T, which maps a semi-ordered Banach space into itself, is called to be positive if $x \geqq 0$ implies $T(x) \geqq 0$.

[3] $(\mathbf{M})$ is the space of all the (real- or complex-valued) completely additive set functions $x(E)$ defined for all Borel set E of Ω. For any $x(E) \in (\mathbf{M})$, its norm is defined by: $\| x \| =$ total variation of $| x(E) |$ on Ω. In $(\mathbf{M})$, a real-valued completely additive set function $x(E)$ is called to be positive (and is denoted by $x \geqq 0$), if $x(E) \geqq 0$ for any Borel set $E \subset \Omega$.

[4] (M^*) is the space of all the (real- or complex-valued) bounded Borel measurable functions $x(t)$ defined on Ω. For any $x(t) \in (M^*)$, its norm is defined by: $\| x \| = \text{l.u.b.}_{t \in \Omega} | x(t) |$. In (M^*), two functions, which differ from each other at least at one point, are considered to be different; it is to noted that the Banach space (M^*) is essentially different from (M), where the norm is defined by: $\| x \| = \text{ess. max.}_{t \in \Omega} | x(t) |$. In (M^*), a real-valued bounded Borel measurable function $x(t)$ is called to be positive if $x(t) \geqq 0$ for any $t \in \Omega$.

into itself. If, in particular, $P(t, E)$ has the density $p(t, s)$, then these operations become

$$x \to T(x) = y: \quad y(E) = \int_E \left(\int_\Omega x(dt) p(t, s) \right) ds,$$

$$x \to \bar{T}(x) = y: \quad y(t) = \int_\Omega p(t, s) x(s)\, ds;$$

consequently, in the former case, T may also be considered as a bounded positive linear operation which maps the Banach space (L)[5] into itself:

$$x \to T(x) = y: \quad y(s) = \int_\Omega x(t)\, p(t, s)\, dt.$$

In all these cases, T and $\bar{T}$ are both of norm 1, and it will be easily seen that the iterates T^n and $\bar{T}^n$ of T and $\bar{T}$ respectively, which are defined by the iterated kernel $P^{(n)}(t, E)$:

$$x \to T^n(x) = y: \quad y(E) = \int_\Omega x(dt)\, P^{(n)}(t, E),$$

$$x \to \bar{T}^n(x) = y: \quad y(t) = \int_\Omega P^{(n)}(t, ds)\, x(s),$$

have also the same properties. Thus our problem is transformed into the investigation of the behavior of the iterates T^n of a bounded positive linear operation T, which maps the Banach space $(\mathbf{M})$, (L) or (M^*) into itself.

N. Kryloff-N. Bogolioùboff [1], [2] introduced the condition that

(K) $\begin{cases} \textit{there exist an integer } m \textit{ and a strongly completely continuous}^6 \textit{ linear} \\ \textit{operation } V, \textit{ which maps } (\mathbf{M}) \textit{ into itself, such that } \| T^m - V \| < 1, \end{cases}$

and under this condition they have obtained a remarkable result. In the present paper we shall develop this idea of N. Kryloff-N. Bogolioùboff and shall obtain a more precise result. This investigation will be carried out in §4, and our principal results may be summed up in the following two statements:

(1) *the condition* (D) *implies the condition* (K),

(2) *under the condition* (K), *all the results of W. Doeblin can be obtained even in a more precise form.*

Besides the *ergodic kernel* (= *ensemble final* due to W. Doeblin) we shall introduce a new notion of *ergodic part*, and it will be shown that this also plays a fundamental rôle in the investigation of the asymptotic behavior of the sequence $P^{(n)}(t, E)$ for large n.

In order to obtain these results, we shall develop in §§2 and 3 a general theory of iteration of bounded linear operations in Banach spaces. Although the

[5] (L) is, as usual, the space of all the (real- or complex-valued) measurable functions $x(t)$ which are absolutely integrable on Ω. For any $x(t) \in (L)$, its norm is defined by: $\| x \| = \int_\Omega | x(t) | \, dt$, and in (L), an integrable function $x(t)$ is called to be positive if we have $x(t) \geqq 0$ almost everywhere in t.

iteration of bounded linear operations in concrete Banach space (for example, in Hilbert space) was discussed by many authors, such a theory has not hitherto been developed in general Banach spaces. We shall show now that this is possible under general conditions, and the results thus obtained will find their full applications in §4.

Thus §§2 and 3 may be considered as preparations to §4. Nevertheless, these are also the principal chapters of this paper. Indeed, Theorem 1 (mean ergodic theorem) and Theorem 4 (uniform ergodic theorem) are among the main results of our paper. In order to explain the meaning of these theorems, let us recall the *mean ergodic theorem* of J. v. Neumann:

Let S be a one-to-one measure-preserving transformation of a space Ω into itself. Then

$$x \rightarrow T(x) = y: \qquad y(t) = x(S(t))$$

is a positive unitary transformation which maps the Hilbert space (L^2) (defined on Ω) into itself, and the mean ergodic theorem of J. v. Neumann says that, for any $x \in (L^2)$, the sequence $\{x_n\}$ $(n = 1, 2, \cdots)$, where $x_n = \frac{1}{n}(T + T^2 \cdots + T^n)x$, converges strongly to some element $\bar{x} \in (L^2)$.

This theorem may be generalized in two ways: firstly, (L^2) may be substituted by an arbitrary Banach space (B), and secondly, T may be substituted by an arbitrary quasi-weakly completely continuous[6] linear operation which maps the Banach space (B) into itself. The positiveness of T and the existence of the inverse operation T^{-1} are both not assumed. This theorem is called the *mean ergodic theorem in Banach spaces* and will be proved in §2.[7]

Moreover, if T is quasi-strongly completely continuous,[6] i.e., if T satisfies the condition (K) of N. Kryloff-N. Bogolioùboff, then the strong convergence in the mean ergodic theorem may be substituted by the uniform one. This is a generalization of the results of M. Fréchet [1], [2], C. Visser [1] and N. Kryloff-N. Bogolioùboff [1], [2], and will be called the *uniform ergodic theorem in Banach spaces*. We shall prove this in §3. This theorem is extremely powerful in the investigation of concrete Markoff's process in §4.

In concluding the introduction, we express our hearty thanks to Yukio Mimura and Shôkichi Iyanaga for their constant encouragements during the preparation of this paper.

Chapter 2. Mean Ergodic Theorem in Banach Spaces[8]

(Generalization of J. v. Neumann's mean ergodic theorem)

In this chapter we are concerned with weakly completely continuous and quasi-weakly completely continuous linear operations in general (real or com-

[6] This notion will be explained at the beginning of §2.

[7] Recently, the generalizations of the mean ergodic theorem were given also by N. Wiener [1], [2], G. Birkhoff [2] and N. Dunford [1]. We shall discuss these results in §2. See the remarks after Theorem 1 of §2.

[8] K. Yosida [2], S. Kakutani [1], K. Yosida and S. Kakutani [1].

plex) Banach spaces. By definition, a bounded linear operation T, which maps a Banach space (B) into itself, is called to be *weakly completely continuous* if it maps the unit sphere $\| x \| \leqq 1$ of (B) on a weakly compact set of (B). More generally, T is called to be *quasi-weakly completely continuous,* if there exist an integer m and a weakly completely continuous linear operation V, which maps (B) into itself, such that $\| T^m - V \| < 1$.

THEOREM 1. *Let T be a bounded linear operation which maps a Banach space (B) into itself. Let us further assume that*

(2.1) *there exists a constant C such that $\| T^n \| \leqq C$ for $n = 1, 2, \cdots$, and that*

(2.2) *for any $x \in (B)$, the sequence $\{x_n\}$ $(n = 1, 2, \cdots)$, where $x_n = \frac{1}{n}(T + T^2 + \cdots + T^n)x$, contains a subsequence which converges weakly to a point $\bar{x} \in (B)$.*

Under these assumptions,

(2.3) *the sequence $\{x_n\}$ $(n = 1, 2, \cdots)$ converges strongly to a point $\bar{x}$,*

and if
we denote by T_1 the operation $x \to \bar{x}$, then T_1 is a bounded linear operation which maps (B) into itself and

$$TT_1 = T_1T = T_1^2 = T_1, \qquad \| T_1 \| \leqq C. \tag{2.4}$$

PROOF. Let x be an arbitrary point of (B). By (2.2), there exists a subsequence $\{x_{n_\nu}\}$ $(\nu = 1, 2, \cdots)$ of $\{x_n\}$ $(n = 1, 2, \cdots)$ which converges weakly to a point $\bar{x}$ of (B). We shall first prove that we have

$$T(\bar{x}) = \bar{x}. \tag{2.5}$$

This is an easy consequence of the relation:

$$\begin{aligned} &\| T(x_n) - x_n \| \\ &= \left\| \frac{1}{n}(T^2 + T^3 + \cdots + T^{n+1})x - \frac{1}{n}(T + T^2 + \cdots + T^n)x \right\| \\ &= \left\| \frac{1}{n}(T^{n+1} - T)x \right\| \leqq \frac{2C}{n} \| x \|, \end{aligned} \tag{2.6}$$

which is valid for $n = 1, 2, \cdots$. Indeed, putting $n = n_\nu$ in (2.6) and making ν tend to ∞, we have (2.5) at once; for $\{x_{n_\nu}\}$ and $\{T(x_{n_\nu})\}$ converge weakly to $\bar{x}$ and $T(\bar{x})$ respectively.

We shall next prove (2.3). For this purpose, decompose x into two parts: $x = \bar{x} + (x - \bar{x})$. Then we have, by (2.5),

$$x_n = \bar{x} + \frac{1}{n}(T + T^2 + \cdots + T^n)(x - \bar{x}), \tag{2.7}$$

and consequently, in order to prove (2.3), we have only to prove that the second term of the right hand side of (2.7) converges strongly to 0 as $n \to \infty$.

Let the range of the linear operation $I - T$ (I is the identical transformation) be R, and denote its closure (in the strong sense) by $\bar{R}$. Since R is a linear subspace of (B), $\bar{R}$ is also closed in the weak sense. Since the sequence

$$\left\{I - \frac{1}{n}(T + T^2 + \cdots + T^n)\right\} x$$
$$= (I - T)\left(I + \frac{n-1}{n} T + \frac{n-2}{n} T^2 + \cdots + \frac{1}{n} T^{n-1}\right) x$$

converges weakly to $x - \bar{x}$, and since the right hand side clearly belongs to R, $x - \bar{x}$ belongs to $\bar{R}$. Consequently, in order to prove (2.3), we have only to prove that

$$\frac{1}{n}(T + T^2 + \cdots + T^m)x \text{ converges strongly to } 0 \tag{2.8}$$

for any $y \in \bar{R}$. This may be performed in two steps:

1st case. $y \in R$. Taking a point $z \in (B)$ such that $y = z - T(z)$, we have

$$\left\|\frac{1}{n}(T + T^2 + \cdots + T^n)y\right\| = \left\|\frac{1}{n}(T + T^2 + \cdots + T^n)(z - T(z))\right\|$$
$$= \left\|\frac{1}{n}(T - T^{n+1})z\right\| \leqq \frac{2C}{n} \| z \| \to 0.$$

2nd case. $y \in \bar{R}$. For any $\epsilon > 0$ there exists a point $y' \in R$ such that $\| y - y' \| < \epsilon$. For this y' we have

$$\left\|\frac{1}{n}(T + T^2 + \cdots + T^n)y\right\| \leqq \left\|\frac{1}{n}(T + T^2 + \cdots + T^n)y'\right\|$$
$$+ \left\|\frac{1}{n}(T + T^2 + \cdots + T^n)(y - y')\right\|$$
$$\leqq \left\|\frac{1}{n}(T + T^2 + \cdots + T^n)y'\right\| + C\epsilon.$$

Since the first term of the right hand side converges strongly to 0 as $n \to \infty$ (by the result obtained in the 1st case), we have

$$\overline{\lim_{n\to\infty}} \left\|\frac{1}{n}(T + T^2 + \cdots + T^n)y\right\| \leqq C\epsilon;$$

and, since $\epsilon > 0$ is arbitrary, we have (2.8). Thus (2.8) is proved for any $y \in \bar{R}$ and hereby (2.3) is completely proved.

Let now T_1 be the transformation $x \to \bar{x}$. T_1 is clearly linear and bounded: $\| T_1 \| \leqq C$. Since (2.5) is always true, we have $TT_1 = T_1$. Thus $T^n T_1 = T_1$ and $\frac{1}{n}(T + T^2 + \cdots + T^n)T_1 = T_1$ for $n = 1, 2, \cdots$, and consequently $T_1^2 = T_1$. In order to prove the relation: $T_1 T = T_1$, we have only to start from the inequality:

$$\left\| \frac{1}{n}(T + T^2 + \cdots + T^n)T(x) - \frac{1}{n}(T + T^2 + \cdots + T^n)x \right\| \leqq \frac{2C}{n} \| x \| \to 0.$$

Since the terms on the left hand side converge strongly to $T_1T(x)$ and $T_1(x)$ respectively, we have $T_1T(x) = T_1(x)$ for any $x \in (B)$.

Thus the proof of Theorem 1 is completed.

REMARK 1. T_1 is a projection operator which maps (B) on the proper space (B_1) of T belonging to the proper value 1.[9] That is, T_1 does not vanish identically if and only if 1 is a proper value of T, and $T_1(x) = x$ if and only if we have $T(x) = x$.

REMARK 2. In virtue of (2.1), (2.2) is surely satisfied if T is weakly completely continuous. Hence

COROLLARY 1. *Let T be a weakly completely continuous linear operation which maps a Banach space (B) into itself. If there exists a constant C such that $\| T^n \| \leqq C$ for $n = 1, 2, \cdots$, then the sequence of operations $\left\{\frac{1}{n}(T + T^2 + \cdots + T^n)\right\}$ $(n = 1, 2, \cdots)$ converges strongly to a bounded linear operation T_1 which maps (B) into itself and which satisfies (2.4); i.e., mean ergodic theorem is valid in (B).*

The conditions of this Corollary are clearly satisfied if (B) is a Hilbert space and if T is a unitary operation in it. (Since Hilbert space itself is locally weakly compact.) Thus Theorem 1 is, in two respects, a generalization of the mean ergodic theorem of J. v. Neumann [1], [2]. (See also E. Hopf [1].) In the first place, our theorem is true in general complex Banach space, while the proof of J. v. Neumann is valid only in the case of a Hilbert space. In the second place, we do not assume that T is isometric, and the existence of the inverse operation T^{-1} is not required. The positiveness of the operation T is also not required. We have only to assume that the conditions (2.1) and (2.2) are satisfied. The first condition (2.1) is always satisfied with $C = 1$ in the case of ergodic theorems and Markoff's processes; and the second one (2.2) is satisfied in every weakly compact Banach space (for example, in regular Banach space and especially in (L^p) $(p > 1)$) for any bounded linear operation T. Consequently, mean ergodic theorem holds true in (L^p) $(p > 1)$ under the trivial condition (2.1). This result was also obtained independently by F. Riesz [1]. On the contrary, in the Banach space $(L) = (L^1)$, the condition (2.2) is not always satisfied for an arbitrary bounded linear operation T. In order that a bounded subset S of (L) be weakly compact, it is (necessary and) sufficient that the functions belonging to this subset S are uniformly integrable, that is, that there exists for any $\epsilon > 0$ a positive number $\delta > 0$ such that $\mathrm{mes}(E) < \delta$ implies

[9] It is to be noted that in general Banach space (B), there does not necessarily exist, for any closed linear subspace (B'), a projection operator (of norm 1 or of finite norm) which maps (B) in (B').

$\int_E |x(t)|\,dt < \epsilon$ for any $x(t) \in S$. This condition is surely satisfied if there exists a function $x_0(t) \in (L)$ such that $|x| \leqq x_0$ (that is, $|x(t)| \leqq x_0(t)$ almost everywhere in t) for any $x(t) \in S$. Hence we have

COROLLARY 2. *Let T be a bounded linear operation which maps the Banach space (L) into itself. If there exists a constant C such that $\|T^n\| \leqq C$ for $n = 1, 2, \cdots$, and if there exists, for any $x(t) \in (L)$, a function $x_0(t) \in (L)$ such that $T^n(x) \leqq x_0 \left(\text{or, more generally, } \frac{1}{n}(T + T^2 + \cdots + T^n)x \leqq x_0\right)$ for $n = 1, 2, \cdots$, then for any $x(t) \in (L)$ the sequence $\left\{\frac{1}{n}(T + T^2 + \cdots + T^n)x\right\}$ $(n = 1, 2, \cdots)$ converges strongly to some function $\bar{x}(t) \in (L)$; i.e., mean ergodic theorem is valid in (L).*

In order to prove this, we have only to notice that there exists also an x_0' such that $T^n(-x) \leqq x_0'$ i.e. $-x_0' \leqq T^n(x)$ for $n = 1, 2, \cdots$.

This theorem was proved by F. Riesz [2] and an analogous theorem was also stated by Garrett Birkhoff [1] for abstract (L)-spaces. The result of G. Birkhoff is more general in formulation, but it was pointed out that in essential the two theorems are equivalent.[10]

Moreover, if T is an integral operator with the probability density kernel $p(t, s)$:

$$x \to T(x) = y:\ y(s) = \int_\Omega x(t)p(t, s)\,dt$$

$$(p(t, s) \geqq 0 \quad \text{for any } t, s; \qquad \int_\Omega p(t, s)\,ds = 1 \quad \text{for any } t),$$

then a (necessary and) sufficient condition that T be weakly completely continuous is that $p(t, s)$ is uniformly (in t) integrable in s, that is, that there exists for any $\epsilon > 0$ a positive number $\delta > 0$ such that $\text{mes}(E) < \delta$ implies $\int_E p(t, s)\,ds < \epsilon$ for any t. This condition was investigated by J. L. Doob [1] without being noticed that the linear operation T becomes weakly completely continuous under this condition. We have once[11] treated this case of the condition of J. L. Doob as an application of the mean ergodic theorem. But, on looking precisely into the detail of the fact, we have found that, in this case, the linear operation T satisfies even the (in some sense stronger) condition (K) of N. Kryloff-N. Bogolioùboff, and that the uniform ergodic theorem is true in this case. (Indeed, the condition (K) of N. Kryloff-N. Bogolioùboff follows from the condition (D) of W. Doeblin, which is weaker than that of J. L. Doob.)[12]

REMARK 3. Recently Garrett Birkhoff [2] proved that in every uniformly

[10] S. Kakutani [2].

[11] K. Yosida and S. Kakutani [1].

[12] K. Yosida [3].

convex Banach space mean ergodic theorem is valid for any bounded linear operation of norm not exceeding unity. This is also a generalization of the mean ergodic theorem of J. v. Neumann, since Hilbert space is uniformly convex. But this result of G Birkhoff is contained in our Theorem 1; for, every uniformly convex Banach space is regular and consequently locally weakly compact.[13] (It is, however, to be remarked that the proof of G. Birkhoff distinguishes itself by its extreme simplicity.)

REMARK 4. We can also prove the theorem of the following type:

COROLLARY 3. *Let S and T be two bounded linear operations which map a Banach space (B) into itself, and assume that these are commutative between themselves: $ST = TS$. If there exists a constant C such that $\| T^n \| \leqq C$ and $\| S^n \| \leqq C$ for $n = 1, 2, \cdots$, and if, for any $x \in (B)$, the sequence $\{x_{mn}\}$ $(m, n = 1, 2, \cdots)$, where*

$$x_{mn} = \frac{1}{mn} (S + S^2 + \cdots + S^m)(T + T^2 + \cdots + T^n)x,$$

contains a subsequence $\{x_{m_\nu n_\nu}\}$ $(m_\nu \to \infty,\ n_\nu \to \infty)$ which converges weakly to a point $\bar{x}$ of (B), then we have $\lim_{m,n\to\infty} x_{mn} = \bar{x}$ strongly for any $x \in (B)$; and if we denote the transformation $x \to \bar{x}$ by U, then U is a bounded linear operation which maps (B) into itself and we have

$$US = SU = UT = TU = U^2 = U \quad \textit{and} \quad \| U \| \leqq C.$$

PROOF. It will be easily seen (exactly as in the proof of Theorem 1) that we have $S(\bar{x}) = T(\bar{x}) = \bar{x}$. Hence we have only to prove that the sequence $\left\{\frac{1}{mn} (S + S^2 + \cdots + S^m)(T + T^2 + \cdots + T^n)(x - \bar{x})\right\}$ converges strongly to 0 as $m \to \infty$ and $n \to \infty$ simultaneously. In order to prove this, denote by R the set of all the points $z \in (B)$ of the form: $z = (x - S(x)) + (y - T(y))$ with $x, y \in (B)$. R is clearly a linear subspace of (B) and $x - \bar{x}$ belongs to its strong ($=$ weak) closure $\bar{R}$; for, we have

$$\begin{aligned}
&\left\{I - \frac{1}{mn} (S + S^2 + \cdots + S^m)(T + T^2 + \cdots + T^n)\right\}x \\
&\qquad = \left\{I - \frac{1}{m} (S + S^2 + \cdots + S^m)\right\}x \\
&\qquad\qquad + \left\{\frac{1}{m} (S + S^2 + \cdots + S^m)\right. \\
&\qquad\qquad\qquad \left. - \frac{1}{mn} (S + S^2 + \cdots + S^m)(T + T^2 + \cdots + T^n)\right\}x
\end{aligned}$$

[13] S. Kakutani [3]. See also D. Milman [1]. B. J. Pettis [1].

$$= (I - S)\left(I + \frac{m-1}{m}S + \frac{m-2}{m}S^2 + \cdots + \frac{1}{m}S^{m-1}\right)x$$
$$+ \frac{1}{m}(S + S^2 + \cdots + S^m)(I - T)$$
$$\cdot\left(I + \frac{n-1}{n}T + \frac{n-2}{n}T^2 + \cdots + \frac{1}{n}T^{n-1}\right)x,$$

and the left hand side converges weakly to $x - \bar{x}$. Consequently, we have only to prove that $\left\{\frac{1}{mn}(S + S^2 + \cdots + S^m)(T + T^2 + \cdots + T^n)z\right\}$ converges strongly to 0 for any $z \in \bar{R}$. This is clear if $z \in R$, and the general case $z \in \bar{R}$ may be treated in exactly the same manner as in the proof of Theorem 1.

Mean ergodic theorem of this type was first discussed by N. Wiener [1], [2]. N. Wiener treated the case when the inverse operators S^{-1} and T^{-1} both exist, and has also obtained the individual ergodic theorem of G. D. Birkhoff's type. Recently N. Dunford [1] has also announced the theorem of the same kind.

THEOREM 2. *Under the same assumptions as in Theorem 1, we have: for any complex number λ of absolute value 1, there exists a bounded linear operation T_λ, which maps (B) into itself, such that*

$$\frac{1}{n}\left(\frac{T}{\lambda} + \frac{T^2}{\lambda^2} + \cdots + \frac{T^n}{\lambda^n}\right) \text{ converges strongly to } T_\lambda, \tag{2.9}$$

$$T_\lambda T = TT_\lambda = \lambda T_\lambda, \qquad T_\lambda^2 = T_\lambda, \qquad \| T_\lambda \| \leqq C, \tag{2.10}$$

$$\lambda \neq \mu \quad \text{implies} \quad T_\lambda T_\mu = 0, \tag{2.11}$$

$$T_\lambda \neq 0 \text{ if and only if } \lambda \text{ is a proper value of } T. \tag{2.12}$$

In this case, T_λ is a projection operator which maps (B) on the proper space (B_λ) of T belonging to the proper value λ, and $T_\lambda(x) = x$ if and only if we have $T(x) = \lambda x$. If we further put $T' = T - \sum_{i=1}^{k} \lambda_i T_{\lambda_i}$ for any system $\{\lambda_1, \lambda_2, \cdots, \lambda_k\}$ of complex numbers of modulus 1 with $\lambda_i \neq \lambda_j$ $(i \neq j)$ and $T_{\lambda_i} \neq 0$, then we have

$$T_{\lambda_i} T' = T' T_{\lambda_i} = 0 \qquad \text{for} \qquad i = 1, 2, \cdots, k, \tag{2.13}$$

$$TT' = T'T = T'^2, \tag{2.14}$$

$$T^n = \sum_{i=1}^{k} \lambda_i^n T_{\lambda_i} + T'^n \quad \text{for } n = 1, 2, \cdots, \tag{2.15}$$

$$\begin{cases} \lambda \text{ is a proper value of } T' \text{ if and only if it is a proper value of } T \text{ and} \\ \lambda \neq \lambda_i \text{ for } i = 1, 2, \cdots, k. \end{cases} \tag{2.16}$$

PROOF. (2.9) is clear from Theorem 1 if we consider T/λ instead of T. From this (2.10) follows at once. To prove (2.11), multiply (2.9) by T_μ from the right. Then we have, by (2.10),

$$\frac{1}{n}\left(\frac{\mu}{\lambda} + \frac{\mu^2}{\lambda^2} + \cdots + \frac{\mu^n}{\lambda^n}\right) T_\mu \quad \text{converges strongly to } T_\lambda T_\mu.$$

Since, in case $\lambda \neq \mu$, the left hand side converges strongly (even uniformly) to a zero operator, we have (2.11). (2.12) is clear from the Remark 1 to Theorem 1. (2.13), (2.14) and (2.15) may also be proved by easy calculations. To prove lastly (2.16), let λ, $|\lambda| \neq 0$, be a proper value of T and assume that $\lambda \neq \lambda_i$ for $i = 1, 2, \cdots, k$. Then there exists a point $x_0 \in (B)$ ($x_0 \neq 0$) such that $T(x_0) = \lambda x_0$. Therefore, multiplying the both sides by T_{λ_i} from the left, we have, by (2.10), $\lambda_i T_{\lambda_i}(x_0) = \lambda T_{\lambda_i}(x_0)$ and consequently (since $\lambda \neq \lambda_i$ we have $T_{\lambda_i}(x_0) = 0$ for $i = 1, 2, \cdots, k$, so that we have $\lambda x_0 = T(x_0) = \sum_{i=1}^{k} \lambda_i T_{\lambda_i}(x_0) + T'(x_0) = T'(x_0)$. Hence λ is a proper value of T'. Conversely, let λ be a proper value of T'. Then there exists a point $x_0 \in (B)$ ($x_0 \neq 0$) such that $T'(x_0) = \lambda x_0$. Since we have $T(x_0) = T\left(\frac{1}{\lambda} T'(x_0)\right) = \frac{1}{\lambda} TT'(x_0) = \frac{1}{\lambda} T'^2(x_0)$ (by (2.14)) $= \lambda x_0$, λ is also a proper value of T. In order to prove that $\lambda \neq \lambda_i$ for $i = 1, 2, \cdots, k$, assume that $\lambda = \lambda_i$ for some i. Then we have

$$\frac{1}{n}\left(\frac{T}{\lambda_i} + \frac{T^2}{\lambda_i^2} + \cdots + \frac{T^n}{\lambda_i^n}\right) x_0 = x_0 \quad \text{for } n = 1, 2, \cdots,$$

and consequently (by (2.9)) $T_{\lambda_i}(x_0) = x_0 \neq 0$, which leads to a contradiction, since $T_{\lambda_i}(x_0) = T_{\lambda_i}\left(\frac{1}{\lambda} T'(x_0)\right) = \frac{1}{\lambda} T_{\lambda_i} T'(x_0) = 0$ by (2.13).

The proof of Theorem 2 is hereby completed.

THEOREM 3. *In Theorems 1 and 2, the condition (2.2) may be substituted by the condition that T is quasi-weakly completely continuous.*

PROOF. We have only to prove that, under the condition (2.1) and the assumption that T is quasi-weakly completely continuous, there exists for any $x \in (B)$ a subsequence $\{x_{n_\nu}\}$ ($\nu = 1, 2, \cdots$) of $\{x_n\}$ ($n = 1, 2, \cdots$) which converges weakly to a point $\bar{x} \in (B)$. In order to prove this, put $T^m = V + D$ with $\|D\| = \alpha < 1$. Then we have

$$T^{pm} = V_p + D^p, \tag{2.17}$$

where $V_p = T^{pm} - (T^m - V)^p$ is weakly completely continuous with $V_1 = V$ and $\|D^p\| \leqq \alpha^p$.[14] Hence, putting

[14] If V and W are weakly completely continuous, then $V + W$, TV and VT are also weakly completely continuous for any bounded linear operation T. Thus the totality of all the weakly completely continuous linear operations constitutes an ideal in the ring of all bounded linear operations. Expanding $T^{pm} - (T^m - V)^p$, the term T^{pm} vanishes and there remain only those terms (finite in number) which contain at least one V-factor. Consequently V_p is weakly completely continuous.

It is to be remarked that the same is also true for strongly completely continuous linear operations.

$$x_{n,m} = \frac{1}{n}(T + T^2 + \cdots + T^m)x$$

for brevity's sake, we have for $n > pm$

$$\begin{aligned} x_n &= \frac{1}{n}(T + T^2 + \cdots + T^n)x \\ &= \frac{1}{n}(T + T^2 + \cdots + T^{pm})x + \frac{1}{n}T^{pm}(T + T^2 + \cdots + T^{n-pm})x \\ &= x_{n,pm} + T^{pm}(x_{n,n-pm}) = x_{n,pm} + V_p(x_{n,n-pm}) + D^p(x_{n,n-pm}). \end{aligned}$$

Since $\| x_{n,n-pm} \| \leqq C \cdot \| x \|$ for any $n > pm$ (by (2.1)), there exists, for any p, a subsequence $\{n_\nu\}$ ($\nu = 1, 2, \cdots$) of $\{n\}$ ($n = 1, 2, \cdots$) such that $\{ V_p(x_{n_\nu, n_\nu - pm})\}$ ($\nu = 1, 2, \cdots$) converges weakly to a point $\bar{x}_p \epsilon (B)$. Consequently we have

$$\overline{\lim_{\nu \to \infty}} | f(x_{n_\nu}) - f(\bar{x}_p) | \leqq \overline{\lim_{\nu \to \infty}} | f(x_{n_\nu, pm}) | + \overline{\lim_{\nu \to \infty}} | f(D^p(x_{n_\nu, n_\nu - pm})) |$$

for any bounded linear functional $f(x)$ defined on (B). Since

$$| f(x_{n_\nu, pm}) | \leqq \| f \| \cdot \| x_{n_\nu, pm} \| \leqq \| f \| \cdot \frac{pm}{n_\nu} \cdot C \| x \| \to 0,$$

and

$$| f(D^p(x_{n_\nu, n_\nu - pm})) | \leqq \| f \| \cdot \| D^p \| \cdot \| x_{n_\nu, n_\nu - pm} \| \leqq \| f \| \cdot \alpha^p \cdot C \cdot \| x \|,$$

we have

$$\overline{\lim_{\nu \to \infty}} | f(x_{n_\nu}) - f(\bar{x}_p) | \leqq \| f \| \cdot \alpha^p \cdot C \| x \| \tag{2.18}$$

for any bounded linear functional $f(x)$ defined on (B).

Applying the diagonal method, we may assume that (2.18) holds for any bounded linear functional $f(x)$ defined on (B) and for $p = 1, 2, \cdots$. Of course, $\bar{x}_p$ may depend of p. Let us now consider the sequence $\{\bar{x}_p\}$ ($p = 1, 2, \cdots$). From (2.18) we have

$$| f(\bar{x}_p) - f(\bar{x}_q) | \leqq (\alpha^p + \alpha^q) \cdot \| f \| \, C \, \| x \|,$$

and since $f(x)$ is an arbitrary bounded linear functional defined on (B), we have $\| \bar{x}_p - \bar{x}_q \| \leqq (\alpha^p + \alpha^q) \cdot C \cdot \| x \|$ for any p and q. Consequently, (since $\alpha < 1$), $\{\bar{x}_p\}$ ($p = 1, 2, \cdots$) is a fundamental sequence in (B). If we put $\bar{x} = \lim_{p \to \infty} \bar{x}_p$, then we have, for any p, (from (2.18))

$$\begin{aligned} \overline{\lim_{\nu \to \infty}} | f(x_{n_\nu}) - f(\bar{x}) | &\leqq \overline{\lim_{\nu \to \infty}} | f(x_{n_\nu}) - f(\bar{x}_p) | + | f(\bar{x}_p) - f(\bar{x}) | \\ &\leqq \| f \| \cdot \alpha^p \cdot C \| x \| + \| f \| \cdot \| \bar{x}_p - \bar{x} \|. \end{aligned}$$

Since p is arbitrary and since the right hand side tends to 0 as $p \to \infty$, we have $\lim_{\nu \to \infty} | f(x_{n_\nu}) - f(\bar{x}) | = 0$. Since $f(x)$ is an arbitrary bounded linear functional defined on (B), we have thus proved that the sequence $\{x_{n_\nu}\}$ $(\nu = 1, 2, \cdots)$ converges weakly to $\bar{x}$. The proof of Theorem 3 is hereby completed.

REMARK. All what we have obtained in Theorems 1 and 2 is also valid for quasi-weakly completely continuous linear operations. We can also prove several propositions concerning the conjugate operators of such bounded linear operations.[15] But we shall not go into the detail, which is not essential in the further discussions of this paper, and we shall proceed to the case of strongly completely continuous and quasi-strongly completely continuous linear operations.

CHAPTER 3. UNIFORM ERGODIC THEOREM IN BANACH SPACES[16]

(Generalization of the theorems of Fréchet-Kryloff-Bogolioùboff)

In this chapter we shall discuss strongly completely continuous and quasi-strongly completely continuous linear operations. By definition, a bounded linear operation T of a Banach space (B) into itself is called to be *strongly completely continuous* if it maps the unit sphere $\| x \| \leqq 1$ of (B) on a strongly compact set of (B). More generally, T is called to be *quasi-strongly completely continuous*, if there exist an integer m and a strongly completely continuous linear operation V, which maps (B) into itself, such that $\| T^m - V \| < 1$.

Since every strongly completely continuous linear operation is a fortiori weakly completely continuous, all what we have obtained in §2 is also valid for the case of strongly completely continuous or quasi-strongly completely continuous linear operations. Moreover, we can show that in the present case, the *strong* convergence in Theorems 1, 2 and 3 may be substituted by the *uniform* one. These are, as will be seen from the proofs below, the consequences of a theorem of F. Riesz and its generalizations to the case of quasi-strongly completely continuous linear operations (Lemmas 3.1 and 3.2).

THEOREM 4. *Let T be a strongly completely continuous or quasi-strongly completely continuous linear operation which maps a Banach space (B) into itself. If there exists a constant C such that $\| T^n \| \leqq C$ for $n = 1, 2, \cdots$, then the proper values λ of T of modulus 1 (if such proper value ever exists) are finite in number and each of them is of finite multiplicity. Let us denote these by $\lambda_1, \lambda_2, \cdots, \lambda_k$. Then there exists a system of strongly completely continuous linear operations $T_{\lambda_1}, T_{\lambda_2}, \cdots, T_{\lambda_k}$ (each of them $\neq 0$), and a strongly completely continuous or quasi-strongly completely continuous linear operation S (which might vanish), such that T^n is decomposed into the form:*

$$(3.1) \qquad T^n = \sum_{i=1}^{k} \lambda_i^n T_{\lambda_i} + S^n, \qquad n = 1, 2, \cdots,$$

[15] K. Yosida and S. Kakutani [1], K. Yosida [4].
[16] K. Yosida [1], [4], S. Kakutani [1].

with

$$(3.2)\qquad \begin{cases} TT_{\lambda_i} = T_{\lambda_i}T = \lambda_i T_{\lambda_i}, & T_{\lambda_i}^2 = T_{\lambda_i}, \quad T_{\lambda_i}T_{\lambda_j} = 0 \quad (i \neq j), \\ T_{\lambda_i}S = ST_{\lambda_i} = 0, & \| T_{\lambda_i} \| \leqq C, \quad i = 1, 2, \cdots, k, \end{cases}$$

and

$$(3.3)\qquad \| S^n \| \leqq \frac{M}{(1+\epsilon)^n}, \qquad n = 1, 2, \cdots,$$

where M and ϵ are positive constants which are independent of n.

REMARK. Theorem 4 is a generalization of the results of C. Visser, M. Fréchet and N. Kryloff-N. Bogolioùboff. C. Visser [1] discussed the iteration of strongly completely continuous linear operations in Hilbert space, and has obtained only strong convergence. M. Fréchet [1], [2] discussed the case of an integral operator T with bounded density kernel $p(t, s)$:

$$x \to T(x) = y: \quad y(s) = \int_\Omega x(t)p(t, s)\, dt,$$

and has obtained a considerably precise result after somewhat long calculations. It was, however, shown by Y. Mimura and the authors (Lemma 4.5 in §4) that, for such integral operation T with bounded density kernel, the second iterate T^2 already becomes strongly completely continuous as an operator which maps the Banach space (L) into itself. Hence the theorem of M. Fréchet may be considered as a special case of Theorem 4. N. Kryloff-N. Bogolioùboff [1], [2] treated the case of an integral operator with general probability kernel $P(t, E)$, which maps the Banach space $(\mathbf{M})$ into itself:

$$x \to T(x) = y: \quad y(E) = \int_\Omega x(dt)P(t, E),$$

and have obtained the same results (they announced these without proof) under the same conditions as in Theorem 4. It is, however, to be noted, that Theorem 4 holds true even for the general case when the Banach space (B) is arbitrary and when the positiveness of the operation T is not assumed. The more detailed discussions of such operations and their iterations, where the condition of positiveness plays its essential rôle, will be fully developed in §4.

In order to prove Theorem 4, we shall first prove two Lemmas which may be considered as generalizations of the results of F. Riesz [1].

LEMMA 3.1. *If T is quasi-strongly completely continuous, then the proper values λ of T do not accumulate to a point not interior to the unit circle $|\lambda| = 1$ of the complex plane.*

PROOF. Take an integer p so large that we have $\alpha^p < \frac{1}{4}$ where $T^m = V + D$, $\| D \| = \alpha < 1$. We have (exactly as we have obtained (2.17) in §2)

$$(3.4)\qquad T^{pm} = V_p + D^p, \qquad \| D^p \| \leqq \alpha^p < \tfrac{1}{4},$$

where V_p is a strongly completely continuous linear operation which maps (B) into itself. From this it will be easily seen that we have only to consider the case when $m = 1$ and $\alpha < \frac{1}{4}$.

In order to prove Lemma 3.1 in this case, let $\{\lambda_n\}$ $(n = 1, 2, \cdots, \lambda_n \neq \lambda_m$ $(n \neq m))$ be a sequence of proper values of T which converges to a point λ_0 of modulus not smaller than 1. We shall deduce a contradiction from these assumptions. For this purpose, let $\{x_n\}$ $(n = 1, 2, \cdots)$ be a proper element of T corresponding to the proper value λ_n : $T(x_n) = \lambda_n x_n$, $x_n \neq 0$, $n = 1, 2, \cdots$. We shall first prove that these x_n are mutually linearly independent. This may be done by mathematical induction. Let $x_1, x_2, \cdots, x_{n-1}$ be already linearly independent and assume that x_n depends linearly on $x_1, x_2, \cdots, x_{n-1}$: $x_n = \sum_{i=1}^{n-1} \alpha_i x_i$. Then we have

$$\sum_{i=1}^{n-1} \alpha_i(\lambda_n - \lambda_i)x_i = \lambda_n x_n - \sum_{i=1}^{n-1} \alpha_i \lambda_i x_i = T\left(x_n - \sum_{i=1}^{n-1} \alpha_i x_i\right) = 0.$$

Since $\{\alpha_i\}$ $(i = 1, 2, \cdots, n - 1)$ do not vanish simultaneously, this is a contradiction to the assumption that $x_1, x_2, \cdots, x_{n-1}$ are mutually linearly independent.

Thus we have proved that x_n $(n = 1, 2, \cdots)$ are mutually linearly independent. Consequently, the $(n - 1)$-dimensional subspace (X_{n-1}) of (B), which is spanned by $x_1, x_2, \cdots, x_{n-1}$, is a true subspace of the linear space (X_n) which is spanned by $x_1, x_2, \cdots, x_n$. Hence, by a well-known theorem of F. Riesz [1], there exists a sequence of points $\{y_n\}$ $(n = 1, 2, \cdots)$ such that $y_n \in (X_n)$, $\| y_n \| = 1$ and $\| y_n - x \| > \frac{1}{2}$ for any $x \in (X_{n-1})$. Since each y_n is of the form: $y_n = \sum_{i=1}^{n} \alpha_i x_i$, we have

$$y_n - T\left(\frac{y_n}{\lambda_n}\right) = \sum_{i=1}^{n} \alpha_i x_i - \sum_{i=1}^{n} \frac{\alpha_i}{\lambda_n} T(x_i) = \sum_{i=1}^{n-1} \alpha_i \left(1 - \frac{\lambda_i}{\lambda_n}\right) x_i \in (X_{n-1})$$

and

$$T\left(\frac{y_m}{\lambda_m}\right) = \sum_{i=1}^{m} \frac{\alpha_i}{\lambda_m} T(x_i) = \sum_{i=1}^{m} \alpha_i \frac{\lambda_i}{\lambda_m} x_i \in (X_m)$$

for any n and m. Consequently, the trivial identity:

$$T\left(\frac{y_n}{\lambda_n}\right) - T\left(\frac{y_m}{\lambda_m}\right) = y_n - \left\{\left(y_n - T\left(\frac{y_n}{\lambda_n}\right)\right) + T\left(\frac{y_m}{\lambda_m}\right)\right\}$$

implies

$$\left\| T\left(\frac{y_n}{\lambda_n}\right) - T\left(\frac{y_m}{\lambda_m}\right) \right\| > \tfrac{1}{2} \quad \text{for} \quad n > m. \tag{3.5}$$

On the other hand, since V is strongly completely continuous and since the sequence $\left\{\frac{y_n}{\lambda_n}\right\}$ $(n = 1, 2, \cdots)$ is uniformly bounded, there exist two sequences of integers $\{n_\nu\}$ and $\{m_\nu\}$ $(n_\nu > m_\nu, \nu = 1, 2, \cdots)$ such that $\lim_{\nu \to \infty} \left\| V\left(\frac{y_{n_\nu}}{\lambda_{n_\nu}}\right) - V\left(\frac{y_{m_\nu}}{\lambda_{m_\nu}}\right) \right\| = 0$. This is, however, a contradiction with (3.5), since we have by assumption

$$\left\|\left\{T\left(\frac{y_{n_\nu}}{\lambda_{n_\nu}}\right) - T\left(\frac{y_{m_\nu}}{\lambda_{m_\nu}}\right)\right\} - \left\{V\left(\frac{y_{n_\nu}}{\lambda_{n_\nu}}\right) - V\left(\frac{y_{m_\nu}}{\lambda_{m_\nu}}\right)\right\}\right\|$$
$$\leqq \left\|T\left(\frac{y_{n_\nu}}{\lambda_{n_\nu}}\right) - V\left(\frac{y_{n_\nu}}{\lambda_{n_\nu}}\right)\right\| + \left\|T\left(\frac{y_{m_\nu}}{\lambda_{m_\nu}}\right) - V\left(\frac{y_{m_\nu}}{\lambda_{m_\nu}}\right)\right\|$$
$$= \left\|D\left(\frac{y_{n_\nu}}{\lambda_{n_\nu}}\right)\right\| + \left\|D\left(\frac{y_{m_\nu}}{\lambda_{m_\nu}}\right)\right\| \leqq \alpha\left(\left\|\frac{y_{n_\nu}}{\lambda_{n_\nu}}\right\| + \left\|\frac{y_{m_\nu}}{\lambda_{m_\nu}}\right\|\right) \to \frac{2\alpha}{|\lambda_0|} < \tfrac{1}{2}.$$

Thus Lemma 3.1 is completely proved.

Lemma 3.2. *If T is quasi-strongly completely continuous, then the proper space (B_λ) of T belonging to a proper value λ of T of modulus 1 is of finite dimension.*

Proof. This may be done in exactly the same manner as in the proof of Lemma 3.1. We may again assume that $m = 1$ and $\alpha < \frac{1}{4}$.

If, for some proper value λ with $|\lambda| = 1$, the proper space (B_λ) is not of finite dimension, then there exists a sequence of points $x_n \in (B_\lambda)$ such that $\| x_n \| = 1$, $T(x_n) = \lambda x_n (n = 1, 2, \cdots)$ and $\| x_n - x_m \| > \frac{1}{2}$ for $n > m$.

On the other hand, since V is strongly completely continuous, there exist two sequences $\{n_\nu\}$ and $\{m_\nu\}$ ($n_\nu > m_\nu$, $\nu = 1, 2, \cdots$) such that $\lim_{\nu\to\infty} \| V(x_{n_\nu}) - V(x_{m_\nu}) \| = 0$. This will, however, lead to a contradiction since we have

$$\begin{aligned} \tfrac{1}{2} &< \| x_{n_\nu} - x_{m_\nu} \| = \| \lambda(x_{n_\nu} - x_{m_\nu}) \| = \| T(x_{n_\nu}) - T(x_{m_\nu}) \| \\ &\leqq \| V(x_{n_\nu}) - V(x_{m_\nu}) \| + \| D(x_{n_\nu}) - D(x_{m_\nu}) \| \\ &\leqq \| V(x_{n_\nu}) - V(x_{m_\nu}) \| + 2\alpha. \end{aligned}$$

The proof of Lemma 3.2 is hereby completed.

Proof of Theorem 4. Now, in order to prove Theorem 4, let us recall that Theorem 3 is valid in our case. From Lemma 3.1, there exists only a finite number of proper values λ of T on the unit circle $|\lambda| = 1$. Let us denote these proper values by $\lambda_1, \lambda_2, \cdots, \lambda_k$. Then there exists (by Theorem 2) for each λ_i a bounded linear operation T_{λ_i}, which maps (B) on the proper space (B_{λ_i}) of T belonging to the proper value λ_i. T_{λ_i} satisfies all the conditions of Theorem 2 and is strongly completely continuous since each (B_{λ_i}) is of finite dimension by Lemma 3.2. Let us put

$$S = T - \sum_{i=1}^{k} \lambda_i T_{\lambda_i} \tag{3.6}$$

and consider the bounded linear operation S thus defined. We shall prove that this S is also quasi-strongly completely continuous and that its iterations S^n are uniformly bounded. This is an easy consequence of the relation $S^n = T^n - \sum_{i=1}^k \lambda_i^n T_{\lambda_i}$, which follows immediately from (3.6) and the result of Theorem 2. Indeed, there exists a constant $C' \equiv C + \sum_{i=1}^k \| T_\lambda \| \leqq (k + 1)C$ such that $\| S^n \| \leqq \| T^n \| + \sum_{i=1}^k \| T_{\lambda_i} \| \leqq C'$ for $n = 1, 2, \cdots$, and also an integer m and a strongly completely continuous linear operation $V' \equiv V - \sum_{i=1}^k \lambda_i^m T_{\lambda_i}$ such that $\| S^m - V' \| = \| T^m - V \| < 1$.

Moreover, S has no proper value of modulus 1, so that in order to complete the proof of Theorem 4, we have only to prove the following

LEMMA 3.3. *If, in addition to the assumptions in Theorem 4, T has no proper value of modulus 1, then there exist two constants M and $\epsilon > 0$ such that*

$$\|T^n\| \leqq \frac{M}{(1+\epsilon)^n} \quad \textit{for} \quad n = 1, 2, \cdots. \tag{3.7}$$

PROOF. It is sufficient to prove the case $m = 1$. For, if Lemma 3.3 is true for the case $m = 1$, then there exist two constants M' and $\epsilon' > 0$ such that $\| T^{mn} \| \leqq M'/(1 + \epsilon')^n$ for $n = 1, 2, \cdots$, and it is easy to deduce from this the existence of two constants M and $\epsilon > 0$ which satisfy (3.7).

Now, in order to prove Lemma 3.3 in the case $m = 1$, put $T = V + D$, where V is strongly completely continuous and $\| D \| = \alpha < 1$. By a well-known result, $I - \lambda D$ (I is the identical transformation) admits a unique inverse $I + \lambda D(\lambda) \equiv I + \sum_{n=1}^{\infty} \lambda^n D^n$ which is regular in $|\lambda| < \frac{1}{\alpha}$, and we have $(I + \lambda D(\lambda))(I - \lambda T) = (I + \lambda D(\lambda))(I - \lambda V - \lambda D) = I - \lambda V - \lambda^2 D(\lambda) V \equiv I - V(\lambda)$, where $V(\lambda) \equiv \lambda V + \lambda^2 D(\lambda) V$ is regular in λ for $|\lambda| < \frac{1}{\alpha}$, and is strongly completely continuous for each λ with $|\lambda| < \frac{1}{\alpha}$. By Lemma 3.1, there exists a positive number $\eta > 0$ such that T has no proper value λ in $1 - \eta < |\lambda| < 1 + \eta$. Put $2\epsilon = \min\left(\frac{1}{\alpha} - 1, \eta\right)$ and consider the domain Δ: $1 - 2\epsilon < |\lambda| < 1 + 2\epsilon$. The equation $(I - V(\lambda))x = 0$ has no non-trivial solution $x \neq 0$ for each $\lambda \in \Delta$. For, if there exists an $x_0 \neq 0$ with $(I - V(\lambda))x_0 = 0$, then we have $(I - \lambda T)x_0 = (I - \lambda D)(I - V(\lambda))x_0 = 0$, and this is a contradiction since T has no proper value in Δ. Consequently, since $V(\lambda)$ is strongly completely continuous for any $\lambda \in \Delta$, there exists, by a theorem of F. Riesz, a unique inverse $I - K(\lambda)$ of $I - V(\lambda)$ for any $\lambda \in \Delta$; and it will be easily seen that $I - \lambda R(\lambda) = (I - K(\lambda))(I + \lambda D(\lambda))$ is an inverse of $I - \lambda T$ for each $\lambda \in \Delta$.

Thus we have proved that $I - \lambda T$ has an inverse for any $\lambda \in \Delta$. Since it is clear from the uniform boundedness of $\{T^n\}$ ($n = 1, 2, \cdots$) that $I - \lambda T$ has an inverse $I + \sum_{n=1}^{\infty} \lambda^n T^n$ for any λ with $|\lambda| < 1$, we have thus proved the existence of an inverse $(I - \lambda T)^{-1}$ for any λ with $|\lambda| < 1 + 2\epsilon$. Consequently, by a theorem of M. Nagumo [1], $(I - \lambda T)^{-1}$ is regular in λ for $|\lambda| < 1 + 2\epsilon$ and the series of C. Neumann: $(I - \lambda T)^{-1} = I + \sum_{n=1}^{\infty} \lambda^n T^n$ converges in the uniform sense in $|\lambda| < 1 + 2\epsilon$. Hence there exists a constant M such that the inequality (3.7) is valid for $n = 1, 2, \cdots$.

The proof of Lemma 3.3 and herewith the proof of Theorem 4 are completed.

COROLLARY. *Under the same assumptions as in Theorem 4 we have:*

(i) *For any complex number λ with $|\lambda| = 1$, there exists a strongly completely continuous linear operation T_λ, which maps (B) into itself, such that*

$$\left\| \frac{1}{n}\left(\frac{T}{\lambda} + \frac{T^2}{\lambda^2} + \cdots + \frac{T^n}{\lambda^n}\right) - T_\lambda \right\| \leqq \frac{M}{n} \quad \text{for} \quad n = 1, 2, \cdots,$$

where M is a constant which is independent of n, and T_λ does not vanish identically if and only if λ is a proper value of T.

(ii) *In order that the sequence $\{T^n\}$ $(n = 1, 2, \cdots)$ converge in the uniform sense to a zero operation, it is necessary and sufficient that T has no proper value of absolute value* 1.

(iii) *In order that the sequence $\{T^n\}$ $(n = 1, 2, \cdots)$ converge in the uniform sense to a bounded linear operation $T_1 \neq 0$, it is necessary and sufficient that* 1 *is a proper value of T and that T has no other proper value of absolute value* 1.

In (ii) *and in* (iii), *if the sequence $\{T^n\}$ $(n = 1, 2, \cdots)$ converges in the uniform sense, then it is of the order of geometrical progression; that is, there exist a constant M and a positive number ϵ both independent of n, such that we have $\| T^n \| \leqq M/(1 + \epsilon)^n$ and $\| T^n - T_1 \| \leqq M/(1 + \epsilon)^n$ respectively for $n = 1, 2, \cdots$.*

CHAPTER 4. MARKOFF'S PROCESS[17]

§4.1. Introduction. Let us denote by $P(t, E)$ the transition probability that a point t of the unit interval $\Omega = (0, 1)$ is transferred, by a simple Markoff's process, into a Borel set E of Ω after the elapse of a unit-time. We have always $P(t, E) \geqq 0$ and $P(t, \Omega) = 1$. We shall assume that $P(t, E)$ is completely additive for Borel sets E if t is fixed, and that $P(t, E)$ is Borel measurable in t if E is fixed. Then the transition probability $P^{(n)}(t, E)$ that a point $t \in \Omega$ is transferred into a Borel set E of Ω after the elapse of n unit-times is given recurrently by

$$\begin{aligned} P^{(n)}(t, E) &= \int_\Omega P^{(n-1)}(t, ds)P(s, E) = \int_\Omega P(t, ds)P^{(n-1)}(s, E), \\ &\qquad\qquad\qquad\qquad\qquad n = 2, 3, \cdots, \\ P^{(1)}(t, E) &= P(t, E), \end{aligned} \tag{4.1}$$

where the integration is of Radon-Stieltjes type.

Consider the complex Banach space $(\mathbf{M})$ of the complex-valued completely additive set functions $x(E)$ defined for all Borel set E of Ω. For any $x(E) \in (\mathbf{M})$, its norm is defined by: $\| x \|$ = total variation of $| x(E) |$ on Ω. Then we have

LEMMA 4.1. *The integral operator*

$$x \to T(x) = y: \quad y(E) = \int_\Omega x(dt)P(t, E) \tag{4.2}$$

is a bounded linear operation which maps the Banach space $(\mathbf{M})$ *into itself and* $\| T \| = 1$.

[17] K. Yosida and S. Kakutani [1], K. Yosida [3], S. Kakutani [4].

On the other hand, if we consider the complex Banach space (M^*) of all the complex-valued bounded Borel measurable functions $x(t)$ defined on Ω, with $\| x \| = \underset{t \in \Omega}{\text{l.u.b.}} \, | x(t) |$ as its norm, then we have

LEMMA 4.2. *The integral operator*

$$x \to \overline{T}(x) = y: \quad y(t) = \int_\Omega P(t, ds)x(s) \tag{4.3}$$

is a bounded linear operation which maps the Banach space (M^*) into itself and $\| T \| = 1$.

LEMMA 4.3. *For any $x(E) \in (\mathbf{M})$ and $y(t) \in (M^*)$, we have*

$$\int_\Omega x(dt) \left(\int_\Omega P(t, ds)y(s) \right) = \int_\Omega x_1(ds)y(s), \tag{4.4}$$

where $x_1(E) = \int_\Omega x(dt)P(t, E)$.

These three Lemmas are clear from the properties of $P(t, E)$.

REMARK. By virtue of Lemmas 4.1 and 4.2, $P^{(n)}(t, E)$ can be defined recurrently by (4.1). Hence $P^{(n)}(t, E)$ is completely additive for Borel sets E if t is fixed, and $P^{(n)}(t, E)$ is Borel measurable in t if E is fixed. Clearly we have

$$P^{(n)}(t, E) \geqq 0, \qquad P^{(n)}(t, \Omega) = 1, \qquad n = 1, 2, \cdots . \tag{4.5}$$

Moreover, by the repeated use of Lemma 4.3, we have

$$P^{(m+n)}(t, E) = \int_\Omega P^{(m)}(t, ds)P^{(n)}(s, E) \tag{4.6}$$

for any m and n, and it will be easily seen that the iterated operators T^n and $\overline{T}^n$ are defined by the kernel $P^{(n)}(t, E)$. We have clearly

$$\| T^n \| = \| \overline{T}^n \| = 1, \qquad n = 1, 2, \cdots . \tag{4.7}$$

It is now the purpose of this chapter to investigate the asymptotic behavior of the sequence $\{P^{(n)}(t, E)\}$ for large n. We shall treat this problem by considering $P^{(n)}(t, E)$ as the kernel of the integral operators T^n and $\overline{T}^n$. Since the Banach spaces $(\mathbf{M})$ and (M^*) are not conjugate to each other, these two operators T^n and $\overline{T}^n$ are not the conjugate operators to each other in the strict sense which was given by S. Banach [1]. But in essential, these play the same rôle.

Our problem is not quite easy if we have no further assumptions on the kernel $P(t, E)$. Our fundamental assumptions are the conditions (D) and (K) which were stated in §1. The first condition (D), which is due to W. Doeblin [1] is more general than those given by B. Hostinsky, M. Fréchet and J. L. Doob. The second one (K) is due to N. Kryloff-N. Bogolioùboff [1], [2] and was introduced by them independently of W. Doeblin. We shall show (§4.7) that the condition (D) implies (K), and under the condition (K) all the results of W. Doeblin will be obtained in a more precise form (§§4.2–4.6). Our principal

results are stated in Theorems 5–12. Theorem 5 is a restatement of the uniform ergodic theorem (Theorem 4 of §3), and this is a starting point of all the discussions of this chapter. Among other theorems, Theorem 6 is to be noticed. The formula (4.23) will show how the notions of the Banach spaces (**M**) and (M^*) are essential in our problems. As a corollary to Theorem 6, we shall obtain the new notion of ergodic parts (Theorem 7); and the decomposition of Ω into ergodic kernels (= "ensembles finals" of W. Doeblin) and the dissipative part (Theorem 8) is also a direct consequence of Theorem 6. Moreover, using the fact that under the condition (K) each proper value of T of modulus 1 is a root of unity (Theorem 9), the subdivision of the ergodic parts (and ergodic kernels) into cyclic parts will be easily obtained (Theorem 11 and its Corollary).

The classical results concerning the Markoff's process with a finite number of possible states may be easily obtained from our Theorems. In order to obtain these results, we have only to take a kernel $P(t, E)$ of the special type. This will be carried out in §4.8. In this way, the hitherto known results concerning the Markoff's process with a finite number or a continuum of possible states are obtained in a more precise form by a unified method.

§4.2. Spectral decomposition of $P^{(n)}(t, E)$ under the condition (K).

Theorem 5. *Under the condition* (K), $P^{(n)}(t, E)$ *is decomposed into the form*:

$$P^{(n)}(t, E) = \sum_{i=1}^{k} \lambda_i^n P_{\lambda_i}(t, E) + S^{(n)}(t, E), \qquad n = 1, 2, \cdots, \tag{4.8}$$

where $\{\lambda_i\}$ $(i = 1, 2, \cdots, k)$ *are the proper values of* T *of modulus* 1, *and*

$$\operatorname*{l.u.b.}_{t \in \Omega, E \subset \Omega} \left| \frac{1}{n} \sum_{m=1}^{n} \frac{P^{(m)}(t, E)}{\lambda_i^m} - P_{\lambda_i}(t, E) \right| \leqq \frac{M}{n}, \tag{4.9}$$

$$\int_\Omega P^{(n)}(t, ds) P_{\lambda_i}(s, E) = \int_\Omega P_{\lambda_i}(t, ds) P^{(n)}(s, E) = \lambda_i^n P_{\lambda_i}(t, E), \tag{4.10}$$

$$\int_\Omega P_{\lambda_i}(t, ds) P_{\lambda_j}(s, E) = P_{\lambda_i}(t, E) \quad \text{or } 0 \text{ according as } i = j \text{ or } i \neq j, \tag{4.11}$$

$$\int_\Omega P_{\lambda_i}(t, ds) S(s, E) = \int_\Omega S(t, ds) P_{\lambda_i}(s, E) = 0, \tag{4.12}$$

$$\operatorname*{l.u.b.}_{t \in \Omega, E \subset \Omega} | S^{(n)}(t, E) | \leqq \frac{M}{(1 + \epsilon)^n}, \tag{4.13}$$

$i = 1, 2, \cdots, k$; $n = 1, 2, \cdots$, *where* M *and* ϵ *are positive constants which are independent of* n.

Proof. This theorem follows directly from Theorem 4 (uniform ergodic theorem) of §3. We have only to notice that, by Theorem 4, only the decomposition of the operators T^n is given and that the decomposition of the kernels

$P^{(n)}(t, E)$ is not yet obtained. But, since the convergence $\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} T^m/\lambda_i^m = T_{\lambda_i}$ is uniform, the decomposition of the kernels is simultaneously obtained.

Remark. The uniform limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \frac{P^{(m)}(t, E)}{\lambda^m} = P_\lambda(t, E)$$

exists for any λ with $|\lambda| = 1$. By putting $\lambda = 1$ and remembering (4.5), we see that $P_1(t, E)$ is not identically zero. Hence $\lambda = 1$ is a proper value of T. If we put $\lambda_1 = 1$, then the integral operator T_1 defined by the kernel

$$(4.14) \qquad P_1(t, E) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(t, E) \qquad \text{(uniform limit)}$$

is a projection operator, which maps the Banach space (**M**) on the proper space of T belonging to the proper value 1. More precisely, we have

$$(4.15) \qquad \left\| \frac{T + T^2 + \cdots + T^n}{n} - T_1 \right\| \leqq \frac{M}{n}, \qquad n = 1, 2, \cdots,$$

with a positive constant M. This is a result of N. Kryloff-N. Bogolioùboff. (The existence of the mean sojourn.) Moreover, the integral operators T_{λ_i} $(i = 1, 2, \cdots, k)$ defined by the kernels $P_{\lambda_i}(t, E)$ are all strongly completely continuous.

§4.3. Structure of the kernel $P_1(t, E)$. By (4.5) and (4.14), we have

$$(4.16) \qquad P_1(t, E) \geqq 0, \qquad P_1(t, \Omega) \equiv 1,$$

and the proper value equation in (**M**):

$$(4.17) \qquad T(x) = x: \quad x(E) = \int_\Omega x(dt)P(t, E)$$

admits a non-trivial solution $x \neq 0$. In fact, by (4.10) of Theorem 5, for any $t_0 \in \Omega$, $P_1(t_0, E)$ is a solution of (4.17):

$$(4.18) \qquad P_1(t_0, E) = \int_\Omega P_1(t_0, dt)P(t, E),$$

and $P_1(t, E)$ is not identically zero.

In this section, we shall study the general form of such a solution of (4.17), and using the results thus obtained, the structure of the kernel $P_1(t, E)$ will be determined.

In order to make our discussions clearer, we shall make use of some elementary notions from the theory of semi-ordered Banach spaces. A completely additive real-valued set function $x(E) \in$ (**M**) is called to be positive and is denoted by $x \geqq 0$, if we have $x(E) \geqq 0$ for any Borel set E of Ω; and for any pair of real-valued set functions $x(E)$, $y(E) \in$ (**M**), denote by $x \geqq y$ the relation $x - y \geqq 0$.

Then the relation $x \geqq y$ determines a semi-ordering of the real Banach space $(\mathbf{M})$.

For any real-valued set function $x(E) \in (\mathbf{M})$, its total variation $\bar{x}(E)$ is a positive element of $(\mathbf{M})$, and if we put

$$x_+(E) = \tfrac{1}{2}(\bar{x}(E) + x(E)), \qquad x_-(E) = \tfrac{1}{2}(\bar{x}(E) - x(E)),$$

then these are also the positive elements of $(\mathbf{M})$. These are called the positive and the negative variations of x (on E) respectively. We have $x = x_+ - x_-$, $\bar{x} = x_+ + x_-$ and, from the trivial relation: $(-x)_+ = x_-$, we have $x - (x - y)_+ = y - (y - x)_+$. This last common element is called the minimum of two elements x and y, and is denoted by $x \wedge y$.[18] The special case $x \wedge y = 0$ requires our special attention. It is to be remarked that for two positive elements x and y of $(\mathbf{M})$, $x \wedge y = 0$ is equivalent to the condition that there exist two disjoint Borel sets E_1 and E_2 of Ω such that $x(E_1) = x(\Omega)$ and $y(E_2) = y(\Omega)$. This fact is needed in the following discussions. It is also to be noted that we have always $x_+ \wedge x_- = 0$, $(x - y)_+ \wedge (y - x)_+ = 0$ and

$$(4.19) \qquad (x - (x \wedge y)) \wedge (y - (x \wedge y)) = 0.$$

LEMMA 4.4. *If x and y are two real-valued solutions of* (4.17), *then $\bar{x}$, x_+, x_- and $x \wedge y$ are also solutions of* (4.17).

PROOF. It is sufficient to prove this for $\bar{x}$. Since $P(t, E) \geqq 0$ for any Borel set $E \subset \Omega$, we have

$$\bar{x}(E) \leqq \int_\Omega \bar{x}(dt)P(t, E),$$

and, since $P(t, \Omega) \equiv 1$, here must stand the equal sign.

LEMMA 4.5. *There exists a system of completely additive set functions $\{x_\alpha(E)\}$ $(\alpha = 1, 2, \cdots, l) \in (\mathbf{M})$, with the properties*:

$$(4.20) \qquad T(x_\alpha) = x_\alpha, \quad x_\alpha \geqq 0, \quad x_\alpha(\Omega) = 1, \quad x_\alpha \wedge x_\beta = 0 \quad (\alpha \neq \beta),$$

such that any $x(E) \in (\mathbf{M})$ which satisfies

$$(4.21) \qquad T(x) = x, \quad x \geqq 0, \quad x(\Omega) = 1,$$

is uniquely expressed as a linear combination:

$$(4.22) \qquad x(E) = \sum_{\alpha=1}^{l} c_\alpha x_\alpha(E), \quad c_\alpha \geqq 0, \sum_{\alpha=1}^{l} c_\alpha = 1.$$

PROOF. Let l be the maximum number of elements $x_1, x_2, \cdots, x_l \in (\mathbf{M})$ which satisfy (4.20). The existence of such an l is clear from the quasi-strong complete continuity of the operation T. For, such $x_1, x_2, \cdots, x_l$ are clearly mutually linearly independent and the proper space of T belonging to the proper value 1 is of finite dimension by Lemma 3.2 of §3.

[18] Indeed, it will be easily seen that $x \wedge y$ is the minimum of x and y in the sense of lattice, that is, $x \wedge y \leqq x$, $x \wedge y \leqq y$, and for any $z \in (M)$ with $z \leqq x$, $z \leqq y$, we have $z \leqq x \wedge y$.

We shall prove that this system $\{x_\alpha(E)\}$ $(\alpha = 1, 2, \cdots, l)$ is just the required one. In order to show this, let $x(E)$ be an arbitrary element of $(\mathbf{M})$ which satisfies (4.21). We shall show that $x(E)$ is a linear combination of $x_1, x_2, \cdots, x_l$. For this purpose, consider the minimum $x'_\alpha = x \wedge x_\alpha$ of x and x_α for $\alpha = 1, 2, \cdots, l$. By Lemma 4.4, x'_α is also a solution of (4.17). We shall first show that each x'_α is a constant multiple of $x_\alpha : x'_\alpha = c_\alpha x_\alpha$. Indeed, if this is not true for some α, then x_α and $x''_\alpha \equiv x'_\alpha / \| x'_\alpha \|$ are not equal, and consequently $(x_\alpha - x''_\alpha)_+$ and $(x''_\alpha - x_\alpha)_+$ are both $\neq 0$. Again by Lemma 4.4, these are also the solutions of (4.17). Hence, if we put $x_{\alpha 1} = (x_\alpha - x''_\alpha)_+ / \| (x_\alpha - x''_\alpha)_+ \|$, $x_{\alpha 2} = (x''_\alpha - x_\alpha)_+ / \| (x''_\alpha - x_\alpha)_+ \|$, then the system of $l + 1$ elements $x_1, x_2, \cdots, x_{\alpha-1}, x_{\alpha 1}, x_{\alpha 2}, x_{\alpha+1}, \cdots, x_l$ clearly satisfies (4.20), and this is a contradiction to the definition of l.

Thus we have proved that each $x'_\alpha = x \wedge x_\alpha$ is expressed in the form: $x'_\alpha = c_\alpha x_\alpha$, where c_α is a real number with $0 \leqq c_\alpha \leqq 1$. Next we shall prove that we have $x = \sum_{\alpha=1}^{l} x'_\alpha \equiv \sum_{\alpha=1}^{l} c_\alpha x_\alpha$. For this purpose, we shall show that $x' = x - \sum_{\alpha=1}^{l} x'_\alpha$ satisfies $x' \wedge x_\alpha = 0$ for $\alpha = 1, 2, \cdots, l$. In order to prove this, it is sufficient to show that we have $(x - x'_\alpha) \wedge x_\alpha = 0$ for $\alpha = 1, 2, \cdots, l$. This is, however, clear if $c_\alpha = 1$; for, $x'_\alpha \equiv x \wedge x_\alpha = x_\alpha$ implies $x = x_\alpha$ (since $x(\Omega) = x_\alpha(\Omega) = 1$), and $x' = x - x'_\alpha = 0$. And, in case $c_\alpha < 1$, this follows from $(1 - c_\alpha)((x - x'_\alpha) \wedge x_\alpha) \leqq (x - x'_\alpha) \wedge (1 - c_\alpha)x_\alpha = (x - x'_\alpha) \wedge (x_\alpha - x'_\alpha) = 0$ (by (4.19)). Thus $x' \wedge x_\alpha = 0$ is proved for $\alpha = 1, 2, \cdots, l$. Consequently, if we have $x' \neq 0$, then the system of $l + 1$ elements $x'/\| x' \|, x_1, x_2, \cdots, x_l$ will again satisfy (4.20), and this is also a contradiction.

Thus we have proved that we have $x' = 0$, and consequently $x = \sum_{\alpha=1}^{l} x'_\alpha = \sum_{\alpha=1}^{l} c_\alpha x_\alpha$. Since the uniqueness of the expression and the condition $\sum_{\alpha=1}^{l} c_\alpha = 1$ are both clear, the proof of Lemma 4.5 is hereby completed.

COROLLARY. *$\{x_\alpha(E)\}$ $(\alpha = 1, 2, \cdots, l)$ is a base of all the solutions of the proper value equation* (4.17); i.e., *any $x \in (\mathbf{M})$ which satisfies* (4.17) *is uniquely expressed as a linear combination of $x_1, x_2, \cdots, x_l$.*

PROOF. Clear from Lemmas 4.4 and 4.5.

THEOREM 6. *$P_1(t, E)$ is expressible in the form*:

$$P_1(t, E) = \sum_{\alpha=1}^{l} y_\alpha(t) x_\alpha(E), \tag{4.23}$$

where $\{x_\alpha(E)\}$ $(\alpha = 1, 2, \cdots, l)$ is the system of completely additive set functions $\in (\mathbf{M})$, which was defined in Lemma 4.5, *and $\{y_\alpha(t)\}$ $(\alpha = 1, 2, \cdots, l)$ is a system of bounded Borel measurable functions $\in (M^*)$, which satisfy*

$$T(y_\alpha) = y_\alpha, \qquad y_\alpha(t) \geqq 0, \qquad \sum_{\alpha=1}^{l} y_\alpha(t) = 1, \tag{4.24}$$

and

$$\int_\Omega x_\alpha(dt) y_\beta(t) = 1 \quad \text{or } 0 \text{ according as } \alpha = \beta \text{ or } \alpha \neq \beta. \tag{4.25}$$

Moreover, $\{y_\alpha(t)\}$ $(\alpha = 1, 2, \cdots, l)$ *is a base of all the solutions of the proper value equation in* (M^*):

$$\overline{T}(y) = y:\ y(t) = \int_\Omega P(t, ds)y(s), \tag{4.26}$$

and any solution of (4.26) *with* $y(t) \geqq 0$ *can be expressed uniquely in the form:*

$$y(t) = \sum_{\alpha=1}^{l} c_\alpha y_\alpha(t) \tag{4.27}$$

with non-negative constants c_α $(\alpha = 1, 2, \cdots, l)$.

PROOF. Since $P_1(t, E)$ is a solution of (4.17) for any $t \in \Omega$, the expression (4.23) follows directly from Lemma 4.5. If we take a Borel set E'_α such that $x_\alpha(E'_\alpha) = 1$ and $x_\beta(E'_\alpha) = 0$ for any $\beta \neq \alpha$ (the existence of such E'_α follows from the fact that $x_\alpha(\Omega) = 1$ and $x_\beta \wedge x_\alpha = 0$ for any $\beta \neq \alpha$), then (4.23) becomes $P_1(t, E'_\alpha) = y_\alpha(t)$. Hence each $y_\alpha(t)$ is a bounded Borel measurable function of t. We shall next prove (4.24). Since the second and the third relation of (4.24) are clear from (4.16), we have only to prove the first one. From (4.10) of Theorem 5 we have

$$\int_\Omega P(t, ds)P_1(s, E) = P_1(t, E),$$

or, by (4.23),

$$\sum_{\alpha=1}^{l}\left(\int_\Omega P(t, ds)y_\alpha(s)\right)x_\alpha(E) = \sum_{\alpha=1}^{l} y_\alpha(t)x_\alpha(E),$$

and, if we put $E = E'_\alpha$, then we have

$$\int_\Omega P(t, ds)y_\alpha(s) = y_\alpha(t).$$

Thus (4.24) is proved. In order to prove (4.25), start from the relation $T(x_\alpha) = x_\alpha$. Since $T(x) = x$ is equivalent to $T_1(x) = x$, we have $T_1(x_\alpha) = x_\alpha$, or

$$\int_\Omega x_\alpha(dt)P_1(t, E) = x_\alpha(E),$$

and, putting $E = E'_\beta$, we have the required relation (4.25).

Thus the first part of the theorem is proved. The second part may be proved as follows: Let $y(t)$ be a solution of (4.26). From (4.26) we have (exactly as in the preceding case)

$$y(t) = \int_\Omega P_1(t, ds)y(s),$$

and, by (4.23),

$$y(t) = \sum_{\alpha=1}^{l} y_\alpha(t)\left(\int_\Omega x_\alpha(ds)y(s)\right) = \sum_{\alpha=1}^{l} c_\alpha y_\alpha(t)$$

with $c_\alpha = \int_\Omega x_\alpha(ds)y(s)$. Since $y(s) \geqq 0$ implies $c_\alpha \geqq 0$, the proof of Theorem 6 is hereby completed.

§4.4. Ergodic decomposition of Ω under the condition (K). Let us consider the sets $\bar{E}_\alpha = E_t[y_\alpha(t) = 1]$, $\alpha = 1, 2, \cdots, l$. By (4.24), these are mutually disjoint Borel sets.

THEOREM 7. *There exists a system of mutually disjoint Borel sets* $\{\bar{E}_\alpha\}$ $(\alpha = 1, 2, \cdots, l)$, *such that*

$$x_\alpha(\bar{E}_\beta) = 1 \text{ or } 0 \text{ according as } \alpha = \beta \text{ or } \alpha \neq \beta, \tag{4.28}$$

$$P(t, \bar{E}_\alpha) = 1, \qquad t \in \bar{E}_\alpha, \tag{4.29}$$

$$\underset{t \in \bar{E}_\alpha, E \subset \Omega}{\text{l.u.b.}} \left| \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(t, E) - x_\alpha(E) \right| \leqq \frac{M}{n}, \qquad n = 1, 2, \cdots, \tag{4.30}$$

where M is a constant which is independent of n.

REMARK. (4.29) means that, for any α, each point $t \in \bar{E}_\alpha$ is transferred by the Markoff's process $P(t, E)$ inside $\bar{E}_\alpha$, and (4.30) means that the uniform limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(t, E) = P_1(t, E)$$

is independent of the initial point $t \in \bar{E}_\alpha$. Because of these properties, $\bar{E}_\alpha$ $(\alpha = 1, 2, \cdots, l)$ will be called the *ergodic parts* of Ω.

PROOF OF THEOREM 7. (4.28) and (4.29) are the consequences of (4.25) and (4.26) respectively. In order to show this, we have only to prove

LEMMA 4.6. *If $x(E)$ is a completely additive real-valued set function ϵ (**M**) such that $x(\Omega) = 1$ and $x(E) \geqq 0$ for any Borel set $E \subset \Omega$, and if $y(t)$ is a bounded Borel measurable real-valued function ϵ (M^*) such that $0 \leqq y(t) \leqq 1$ for any $t \in \Omega$, then*

$$\int_\Omega x(dt)y(t) = 1$$

implies

$$x(E_0) = 1, \qquad \text{where } E_0 = E_t[y(t) = 1].$$

PROOF. If we put

$$E_n = E_t\left[1 - \frac{1}{n} \leqq y(t) < 1 - \frac{1}{n+1}\right],$$

then we have $\Omega = E_0 + \sum_{n=1}^{\infty} E_n$ and

$$\begin{aligned} 1 = \int_\Omega x(dt)y(t) &= \int_{E_0} x(dt)y(t) + \sum_{n=1}^{\infty} \int_{E_n} x(dt)y(t) \\ &\leqq x(E_0) + \sum_{n=1}^{\infty}\left(1 - \frac{1}{n+1}\right)x(E_n) \\ &\leqq x(E_0) + \sum_{n=1}^{\infty} x(E_n) = x(\Omega) = 1. \end{aligned}$$

Since the equality holds only if we have $x(E_n) = 0$ for $n = 1, 2, \cdots$, we have $x(E_0) = 1$. Thus Lemma 4.6 and hereby (4.28) and (4.29) are proved.

(4.30) is a restatement of the relation (4.9) of Theorem 5 ($i = 1, \lambda_1 = 1$):

$$\underset{t \in \Omega, E \subset \Omega}{\text{l.u.b.}} \left| \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(t, E) - P_1(t, E) \right| \leqq \frac{M}{n}, \qquad n = 1, 2, \cdots,$$

if we only observe that we have

$$P_1(t, E) = x_\alpha(E) \quad \text{for} \quad t \in \bar{E}_\alpha .$$

As is easily seen, $\bar{E}_\alpha$ is not necessarily the set of the smallest measure with the property (4.28). Indeed, there might exist a Borel set $E \subset \bar{E}_\alpha$ such that mes $(E) <$ mes $(\bar{E}_\alpha)$ and $x_\alpha(E) = 1$. If we denote by $E_\alpha(\subset \bar{E}_\alpha)$ the Borel set of the smallest measure among those which satisfy (4.28), then E_α is determined up to a set of measure zero and $E \subset E_\alpha$, mes $(E) > 0$ imply $x_\alpha(E) > 0$. We shall show that, if we take suitably the sets $E_\alpha \subset \bar{E}_\alpha$ $(\alpha = 1, 2, \cdots, l)$, then the following theorem is true:

THEOREM 8. *The Borel sets E_α $(\alpha = 1, 2, \cdots, l)$ and $\Delta = \Omega - \sum_{\alpha=1}^{l} E_\alpha$ satisfy*

$$P_1(t, E_\alpha) = 1, \qquad t \in \bar{E}_\alpha, \tag{4.31}$$

$$P(t, E_\alpha) = 1, \qquad t \in E_\alpha, \tag{4.32}$$

$$\begin{cases} \textit{for any } t \in \bar{E}_\alpha \textit{ and } E \subset E_\alpha, \textit{ mes } (E) > 0 \textit{ implies} \\ P_1(t, E) > 0 \textit{ and consequently there exists a posi-} \\ \textit{tive integer } n = n(t, E) \textit{ such that } P^{(n)}(t, E) > 0, \end{cases} \tag{4.33}$$

$$\underset{t \in \Omega}{\text{l.u.b.}} \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(t, \Delta) \leqq \frac{M}{n}, \qquad n = 1, 2, \cdots, \tag{4.34}$$

where M is a constant which is independent of n.

REMARK. (4.31) means that, for any α, each point $t \in \bar{E}_\alpha$ is transferred finally into E_α (in the sense of arithmetic mean), and (4.32) means that each point $t \in E_\alpha$ is transferred inside E_α. Moreover, by (4.33), E_α is indecomposable into two sets with the property (4.32). In Theorem 12, we shall obtain a more precise result than (4.34):

$$\underset{t \in \Omega}{\text{l.u.b.}} P^{(n)}(t, \Delta) \leqq \frac{M}{(1+\epsilon)^n}, \qquad n = 1, 2, \cdots,$$

with positive constants M and ϵ.

Because of these properties, E_α ($\alpha = 1, 2, \cdots, l$) and Ω will be called the *ergodic kernels* and the *dissipative part* of Ω respectively. It is to be noted that the ergodic kernels are only determined up to a set of measure zero (even under the condition (4.32)), while the ergodic parts $\bar{E}_\alpha$ are strictly determined by $\bar{E}_\alpha = \underset{t}{E}[y(t) = 1]$.

PROOF OF THEOREM 8. Let $E_\alpha^0 (\subset \bar{E}_\alpha)$ be the Borel set of the smallest measure among those which satisfy (4.28). E_α^0 clearly satisfies (4.31) and (4.33) (since $P_1(t, E) = x_\alpha(E)$ for $t \in \bar{E}_\alpha$), but (4.32) is not always automatically satisfied for any $t \in E_\alpha^0$. In order to obtain the required Borel set E_α, we shall construct a sequence of Borel sets $E_\alpha^0 \supset E_\alpha^1 \supset E_\alpha^2 \cdots \supset E_\alpha^n \supset \cdots$ by mathematical induction. Let E_α^n be already defined and assume that we have $x_\alpha(E_\alpha^n) = 1$. Then we define E_α^{n+1} as the set of all $t \in E_\alpha^n$ which satisfies $P(t, E_\alpha^n) = 1$. Since

$$\int_{E_\alpha^n} x_\alpha(dt)P(t, E_\alpha^n) = \int_\Omega x_\alpha(dt)P(t, E_\alpha^n) = x_\alpha(E_\alpha^n) = 1,$$

we have (by Lemma 4.6) $x_\alpha(E_\alpha^{n+1}) = 1$. If we now consider the set $E_\alpha^\infty = \prod_{n=1}^\infty E_\alpha^n$, then $E_\alpha = E_\alpha^\infty$ is the required set. For, we have

$$x_\alpha(E_\alpha^\infty) = \lim_{n\to\infty} x_\alpha(E_\alpha^n) = 1$$

and

$$P(t, E_\alpha^\infty) = \lim_{n\to\infty} P(t, E_0^n) = 1 \quad \text{for } t \in E_\alpha^\infty.$$

Thus we have proved the existence of the Borel set $E_\alpha \subset \bar{E}_\alpha$ which satisfies (4.31), (4.32) and (4.33). Since (4.34) is clear from

$$P(t, \Delta) = \sum_{\alpha=1}^{l} y_\alpha(t)x_\alpha(\Delta) = 0, \qquad t \in \Omega,$$

the proof of Theorem 8 is completed.

§4.5. Proper values of modulus 1 of the operator T.

THEOREM 9. *Under the condition* (K), *each proper value of modulus* 1 *of the operation T is a root of unity.*

PROOF. Let λ, $|\lambda| = 1$, be a proper value of the bounded linear operation T, which maps (**M**) into itself. By Theorem 5, λ may also be considered as a proper value of the bounded linear operation $\overline{T}$, which maps (M^*) into itself. Indeed, (since $P_\lambda(t, E)$ is not identically zero) if we put $x(t) \equiv P_\lambda(t, E)$ for a suitable Borel set E, then $x(t)$ is a non-trivial solution of the proper value equation:

$$(4.35) \qquad T(x) = \lambda x \colon \lambda x(t) = \int_\Omega P(t, ds)x(s).$$

$x(t)$ is clearly a complex-valued bounded Borel measurable function $\epsilon (M^*)$. We shall prove that λ is a root of unity.

In order to prove this, we shall first show that the real-valued bounded measurable function $\bar{x}(t) \equiv |x(t)| \epsilon (M^*)$ attains its maximum. From (4.35), we have

$$\bar{x}(t) \leqq \int_\Omega P(t, ds)\bar{x}(s)$$

and consequently

$$\bar{x}(t) \leqq \int_\Omega P^{(n)}(t, ds)\bar{x}(s), \qquad n = 1, 2, \cdots.$$

Taking the arithmetic mean and considering its limit, we have

$$\bar{x}(t) \leqq \int_\Omega P_1(t, ds)\bar{x}(s)$$

or, by (4.23),

$$\bar{x}(t) \leqq \sum_{\alpha=1}^{l} y_\alpha(t)\left(\int_\Omega x_\alpha(ds)x(s)\right) \equiv \sum_{\alpha=1}^{l} \xi_\alpha y_\alpha(t),$$

where $\xi_\alpha = \int_\Omega x_\alpha(ds)x(s)$ is a real non-negative number. Let ξ be the maximum of ξ_α ($\alpha = 1, 2, \cdots, l$). Since $\sum_{\alpha=1}^{l} y_\alpha(t) \equiv 1$ by (4.24), we have $\bar{x}(t) \leqq \xi$ for any $t \epsilon \Omega$. We shall prove that this ξ is attained by $\bar{x}(t)$ at some point $t_0 \epsilon \Omega$. Indeed, by the definition of ξ, there exists at least one integer α such that $\xi = \int_\Omega x_\alpha(ds)x(s)$, and since $x_\alpha(E) \geqq 0$, $x_\alpha(\Omega) = 1$ and $\bar{x}(s) \leqq \xi$ for any $s \epsilon \Omega$, there must exist, by Lemma 4.6, a point $t_0 \epsilon \Omega$ such that $\bar{x}(t_0) = \xi$ (or more precisely, the set $E_0 = \underset{s}{E}[\bar{x}(s) = \xi]$ satisfies $x_\alpha(E_0) = 1$).

We have thus proved that there exists a point $t_0 \epsilon \Omega$, such that $\bar{x}(t_0) = |x(t_0)| = \max_{t \epsilon \Omega} |x(t)| = \xi$. We shall next prove that the set

$$E(n) = \underset{t}{E}[x(t) = \lambda^n x(t_0)] \tag{4.36}$$

satisfies

$$P^{(n)}(t, E(n)) = 1 \tag{4.37}$$

for $n = 1, 2, \cdots$. From (4.35), we have

$$\lambda^n x(t_0) = \int_\Omega P^{(n)}(t_0, ds)x(s)$$

and, dividing by $\lambda^n x(t_0)$ and taking the real part,

$$1 = \int_\Omega P^{(n)}(t_0, ds)R\left(\frac{x(s)}{\lambda^n x(t_0)}\right), \qquad n = 1, 2, \cdots.$$

Since $P^{(n)}(t_0, E) \geqq 0$, $P^{(n)}(t_0, \Omega) \equiv 1$ and $R\left(\frac{x(s)}{\lambda^n x(t_0)}\right) \leqq 1$ for any $s \in \Omega$, we have (4.37) by Lemma 4.6. For, we have

$$\underset{s}{E}\left[R\left(\frac{x(s)}{\lambda^n x(t_0)}\right) = 1\right] = \underset{t}{E}\,[x(t) = \lambda^n x(t_0)].$$

Thus we have proved (4.37) for $n = 1, 2, \cdots$. From this follows the existence of two positive integers m and n, such that $E(m)\cdot E(n) \neq 0$. For, if we have $E(m)\cdot E(n) = 0$ for any couple of integers m, n $(m \neq n)$, then we have

$$P^{(m)}(t_0, E(l)) = 0 \quad \text{for} \quad m \neq l, \tag{4.38}$$

and consequently, by (4.37),

$$P_1(t_0, E(l)) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(t_0, E(l)) = 0.$$

Since $P_1(t_0, E)$ is completely additive for Borel sets E, this implies

$$P_1(t_0, \sum_{l=1}^{\infty} E(l)) = 0.$$

This is, however, a contradiction, since we have, by (4.37) and (4.38),

$$P^{(m)}(t_0, \sum_{l=1}^{\infty} E(l)) = 1, \qquad m = 1, 2, \cdots,$$

and consequently

$$P_1(t_0, \sum_{l=1}^{\infty} E(l)) = \lim_{m\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{(m)}(t_0, \sum_{l=1}^{\infty} E(l)) = 1.$$

Thus we have $E(m)\cdot E(n) \neq 0$ for a certain couple of integers m, n $(m \neq n)$. Consequently we have (by (4.36)) $\lambda^m x(t_0) = \lambda^n x(t_0) (\neq 0)$ or $\lambda^{m-n} = 1$, and hereby the proof of Theorem 9 is completed.

§4.6. Decomposition of each ergodic part (and ergodic kernel) into subergodic parts (and subergodic kernels).

THEOREM 10. *Under the condition* (K), *there exists a positive integer N such that $P^{(nN)}(t, E)$ is decomposed into the form:*

$$P^{(nN)}(t, E) = P_1^*(t, E) + S^{*(n)}(t, E), \qquad n = 1, 2, \cdots, \tag{4.39}$$

in such a way that we have

$$\begin{aligned} \int_\Omega P^{(N)}(t, ds) P_1^*(s, E) &= \int_\Omega P_1^*(t, ds) P^{(N)}(s, E) \\ &= \int_\Omega P_1^*(t, ds) P_1^*(s, E) = P_1^*(t, E), \end{aligned} \tag{4.40}$$

(4.41) $$\int_\Omega P_1^*(t, ds)S^*(s, E) = \int_\Omega S^*(t, ds)P_1^*(s, E) = 0$$

and

(4.42) $$\underset{t \in \Omega, E \subset \Omega}{\text{l.u.b.}} | S^{*(n)}(t, E) | \leqq \frac{M}{(1 + \epsilon)^n}, \qquad n = 1, 2, \cdots,$$

where M and ϵ are positive constants which are independent of n.

PROOF. In Theorem 9, we have seen that each proper value λ of modulus 1 of T is a root of unity. Since T has only a finite number of such proper values, there exists a (sufficient large) positive integer N such that $\lambda^N = 1$ for any proper value λ of modulus 1 of T. Hence the bounded linear operation T^N corresponding to the kernel $P^{(N)}(t, E)$ has no proper values of modulus 1 other than 1, and Theorem 10 is a direct consequence of Theorem 5.

REMARK. It is to be noted that we have, by (4.39) and (4.42),

(4.43) $$\underset{t \in \Omega, E \subset \Omega}{\text{l.u.b.}} | P^{(nN)}(t, E) - P_1^*(t, E) | \leqq \frac{M}{(1 + \epsilon)^n}, \qquad n = 1, 2, \cdots$$

with positive constants M and ϵ.

Moreover, just as in Lemma 4.5, there exists a system of real-valued completely additive set functions $\{x_i^*(E)\}$ $(i = 1, 2, \cdots, L) \in (\mathbf{M})$ with the properties:

(4.44) $$T^N(x_i^*) = x_i^*, \quad x_i^* \geqq 0, \quad x_i^*(\Omega) = 1, \quad x_i^* \wedge x_j^* = 0 \qquad (i \neq j),$$

such that each $x^*(E) \in (\mathbf{M})$ which satisfies

(4.45) $$T^N(x^*) = x^*, \qquad x^* \geqq 0, \qquad x^*(\Omega) = 1$$

is uniquely expressed in the form:

(4.46) $$x^*(E) = \sum_{i=1}^{L} c_i^* x_i^*(E), \qquad c_i^* \geqq 0, \qquad \sum_{i=1}^{L} c_i^* = 1.$$

In particular, just as in Theorem 6, $P_1^*(t, E)$ is expressed in the form:

(4.47) $$P_1^*(t, E) = \sum_{i=1}^{L} y_i^*(t) x_i^*(E),$$

where $\{y_i^*(t)\}$ $(i = 1, 2, \cdots, L)$ is a system of real-valued bounded Borel measurable functions $\epsilon\ (M^*)$ which satisfy

(4.48) $$T^N(y_i^*) = y_i^*, \qquad y_i^*(t) \geqq 0, \qquad \sum_{i=1}^{L} y_i^*(t) \equiv 1.$$

Let us denote by E_i^* the set $\underset{t}{E}[y^*(t) = 1]$, $i = 1, 2, \cdots, L$. These will be called the *subergodic parts* of Ω. Exactly as in Theorem 7, these are mutually disjoint Borel sets, and we have

$$(4.49)\quad \begin{cases} x_i^*(\bar{E}_j^*) = 1 \text{ or } 0 \text{ according as } i = j \text{ or } i \neq j, \\ P^{(N)}(t, \bar{E}_i^*) = 1, \qquad t \in \bar{E}_i^*, \\ \underset{t \in \bar{E}_i^*, E \subset \Omega}{\text{l.u.b.}} \mid P^{(nN)}(t, E) - x_i^*(E) \mid \leqq \dfrac{M}{(1+\epsilon)^n}, \qquad n = 1, 2, \cdots \end{cases}$$

with positive constants M and ϵ. More precisely we can prove

THEOREM 11. *The totality of all the subergodic parts* $\bar{E}_i^*$ $(i = 1, 2, \cdots, L)$ *is divided into* l *classes* $(\bar{E}_{\alpha_1}^*, \bar{E}_{\alpha_2}^*, \cdots, \bar{E}_{\alpha_{d_\alpha}}^*)$ $(\alpha = 1, 2, \cdots, l)$, *where* d_α *is a divisor of* N *with* $\sum_{\alpha=1}^{l} d_\alpha = L$, *in such a way that*

$$(4.50)\qquad P(t, \bar{E}_{\alpha_{i+1}}^*) = 1, \qquad t \in \bar{E}_{\alpha_i}^*, \qquad i = 1, 2, \cdots, d_\alpha (\alpha_{d_\alpha+1} = \alpha_1),$$

$$(4.51)\qquad \underset{t \in \bar{E}_{\alpha_i}^*, E \subset \Omega}{\text{l.u.b.}} \mid P^{(nd_\alpha)}(t, E) - x_{\alpha_i}^*(E) \mid \leqq \frac{M}{(1+\epsilon)^n}, \qquad n = 1, 2, \cdots,$$

where M *and* ϵ *are positive constants which are independent of* n.

Moreover, all $\bar{E}_{\alpha_i}^*$ $(i = 1, 2, \cdots, d_\alpha)$ *belonging to the same class are contained in the same ergodic part* $\bar{E}_\alpha$, *and if we denote by* $x_{\alpha_i}^*(E)$ *and* $y_{\alpha_i}^*(t)$ *the corresponding elements of* $(\mathbf{M})$ *and* (M^*) *respectively (which are obtained by the arguments given above), then we have*

$$(4.52)\qquad T(x_{\alpha_i}^*) = x_{\alpha_{i+1}}^*, \qquad i = 1, 2, \cdots, d_\alpha(\alpha_{d_\alpha+1} = \alpha_1),$$

$$(4.53)\qquad \bar{T}(y_{\alpha_{i+1}}^*) = y_{\alpha_i}^*, \qquad i = 1, 2, \cdots, d_\alpha(\alpha_{d_\alpha+1} = \alpha_1),$$

$$(4.54)\qquad x_\alpha(E) = \frac{1}{d_\alpha} \sum_{i=1}^{d_\alpha} x_{\alpha_i}^*(E),$$

$$(4.55)\qquad y_\alpha(t) = \sum_{i=1}^{d_\alpha} y_{\alpha_i}^*(t).$$

REMARK 1. (4.50) means that, for any α, each point $t \in \sum_{i=1}^{d_\alpha} \bar{E}_{\alpha_i}^*$ is transferred, by the Markoff's process $P(t, E)$, cyclically in $\bar{E}_{\alpha_1}^*, \bar{E}_{\alpha_2}^*, \cdots, \bar{E}_{\alpha_{d_\alpha}}^*$; and (4.51) means that, in each $\bar{E}_{\alpha_i}^*$, $P^{(d_\alpha)}(t, E)$ defines a Markoff's process, whose n^{th} iterate $P^{(nd_\alpha)}(t, E)$ is uniformly convergent to a limit which is independent of the initial point $t \in \bar{E}_{\alpha_i}^*$.

REMARK 2. The equality $\bar{E}_\alpha = \sum_{i=1}^{d_\alpha} \bar{E}_{\alpha_i}^*$ is not necessarily true. Indeed, $D_\alpha = \bar{E}_\alpha - \sum_{i=1}^{d_\alpha} \bar{E}_{\alpha_i}^*$ is the set of all $t \in \Omega$ such that $y_\alpha(t) \equiv \sum_{i=1}^{d_\alpha} y_{\alpha_i}^*(t) = 1$ and $y_{\alpha_i}^*(t) < 1$ for $i = 1, 2, \cdots, d_\alpha$. From the proof of (4.51), we see that

$$(4.51')\qquad \underset{t \in E_\alpha, E \subset \Omega}{\text{l.u.b.}} \mid P^{(nd_\alpha)}(t, E) - \sum_{i=1}^{d_\alpha} y_{\alpha_i}^*(t) x_{\alpha_i}^*(E) \mid \leqq \frac{M}{(1+\epsilon)^n}, \quad n = 1, 2, \cdots$$

with positive constants M and ϵ.

PROOF OF THEOREM 11. We begin with some preliminary considerations. Since $T^N T(x_i^*) = TT^N(x_i^*) = T(x_i^*)$, $T(x_i^*)$ satisfies (4.43) for $i = 1, 2, \cdots, L$. Hence there exists a system of real constants c_{ij}, $i, j = 1, 2, \cdots, L$, such that

$$(4.56)\qquad T(x_i^*) = \sum_{j=1}^{L} c_{ij} x_j^*, \qquad c_{ij} \geqq 0, \qquad \sum_{j=1}^{L} c_{ij} = 1.$$

Thus T may be considered as a linear transformation on the system $\{x_i^*(E)\}$ $(i = 1, 2, \cdots, L)$. Consider the matrix $C = (c_{ij})$, $i, j = 1, 2, \cdots, L$. This matrix clearly satisfies

$$(4.57)\qquad C^N = \text{unit matrix}.$$

We shall show that C is a matrix composed of only 0 and 1, which defines a permutation of the indices $1, 2, \cdots, L$. Indeed, denoting by $c_{ij}^{(N-1)}$ the ij-elements of the matrix $C^{(N-1)}$, we have $\sum_{k=1}^{L} c_{ik}^{(N-1)} = 1$, $\sum_{k=1}^{L} c_{ik}^{(N-1)} c_{ki} = 1$ for $i = 1, 2, \cdots, L$. Since $0 \leqq c_{ki} \leqq 1$ for any k and i, there must exist, for each i, an index k_i such that $c_{k_i i} = 1$. In other words, each column of the matrix C must contain 1 at least once. Since $\sum_{i=1}^{L} c_{ki} = 1$ for each k, we must have $k_i \neq k_j$ for $i \neq j$. Hence, $(k_1, k_2, \cdots, k_L)$ is a permutation of the indices $(1, 2, \cdots, L)$, and consequently we have $c_{ki} = 0$ for $k \neq k_i$.

Thus we have proved that C defines a permutation of the indices $1, 2, \cdots, L$. Hence the indices $1, 2, \cdots, L$ are divided into $l'(\leqq L)$ classes, and each class is permuted cyclically inside itself by the matrix C. Let these classes be K_α $(\alpha = 1, 2, \cdots, l')$ and the number of the indices belonging to K_α be d_α. By (4.57), each d_α is a divisor of N.

For each α, consider the set of all the indices which belong to K_α. By a suitable numbering $\alpha_1, \alpha_2, \cdots, \alpha_{d_\alpha}$ of these indices, we must have $c_{\alpha_i \alpha_{i+1}} = 1$ and $c_{\alpha_i \alpha_j} = 0$ for $j \neq i + 1$ $(i = 1, 2, \cdots, d_\alpha$; once for all, we put $\alpha_{d_\alpha+1} = \alpha_1)$. Consequently, we have (4.52).

After these preliminaries, we shall proceed to the proof of Theorem 11. We shall first prove that we have $l = l'$ and that there is a one-to-one correspondence between the ergodic parts $\bar{E}_\alpha$ and the classes K_α in such a way that (4.52), (4.53), (4.54) and (4.55) are true. ((4.52) is already proved.)

For this purpose, recall that each $x_\alpha(E)$ satisfies $T(x_\alpha) = x_\alpha$. Hence $x_\alpha(E)$ satisfies (4.45), and consequently $x_\alpha(E)$ is uniquely expressed in the form:

$$(4.58)\qquad x_\alpha(E) = \sum_{i=1}^{L} c_i^* x_i^*(E) \equiv \sum_{\alpha=1}^{l'} \sum_{i=1}^{d_\alpha} c_{\alpha_i}^* x_{\alpha_i}^*(E).$$

We shall first prove that there exists a class K_α such that $c_i^* = 1/d_\alpha$ for $i \in K_\alpha$ and $c_i^* = 0$ for $i \bar{\in} K_\alpha$. In the first place, it is clear that c_i^* is independent of i in each class K_α. For, since $T(x_\alpha) = x_\alpha$ and $T(x_{\alpha_i}) = x_{\alpha_{i+1}}$, we have, from (4.58),

$$x_\alpha(E) = \sum_{\alpha=1}^{l} \sum_{i=1}^{d_\alpha} c_{\alpha_i}^* x_{\alpha_{i+1}}^*(E),$$

and this implies $c_{\alpha_i}^* = c_{\alpha_{i+1}}^*$ for $i = 1, 2, \cdots, d_\alpha$. In the second place, all indices i with $c_i^* > 0$ belong to the same class K_α. For, if this is not the case, then $x_\alpha(E)$ will be decomposed into two non-trivial parts x_α' and x_α'', which are

both invariant under T: $x_\alpha = x'_\alpha + x''_\alpha$, $x'_\alpha > 0$, $x''_\alpha > 0$, $x'_\alpha \wedge x''_\alpha = 0$, $T(x'_\alpha) = x'_\alpha$, $T(x''_\alpha) = x''_\alpha$. If we put $x_{\alpha 1} = x'_\alpha/\| x'_\alpha \|$ and $x_{\alpha 2} = x''_\alpha/\| x''_\alpha \|$, then the system of $l + 1$ elements $x_1, x_2, \cdots, x_{\alpha-1}, x_{\alpha 1}, x_{\alpha 2}, x_{\alpha+1}, \cdots, x_l$ will satisfy the condition (4.20), and this is a contradiction to the definition of l. Thus there must exist a class K_α such that $c_i^* = c > 0$ for $i \in K_\alpha$ and $c_i^* = 0$ for $i \bar{\in} K_\alpha$. Since it is clear that we have $c = 1/d_\alpha$, the relation (4.54) is hereby proved.

Thus to each $x_\alpha(E)$ $(\alpha = 1, 2, \cdots, l)$ there corresponds a class of indices K_α. Conversely, as is easily seen, to each class K_α there corresponds a completely additive set function $x_\alpha(E) = \dfrac{1}{d_\alpha} \sum_{i=1}^{d_\alpha} x^*_{\alpha_i}(E)$ (which clearly belongs to the system determined in Lemma 4.5) in such a way that the correspondence $x_\alpha(E) \leftrightarrow K_\alpha$ is one-to-one. Hence we must have $l = l'$, and the one-to-one correspondence between the ergodic part $\bar{E}_\alpha$ and the class of indices K_α is also established.

We shall next prove (4.55). From the relation

$$\frac{1}{N} \sum_{m=1}^{N} \int_\Omega P_1^*(t, ds) P^{(m)}(s, E) = P_1(t, E)$$

we have, by (4.23) and (4.47),

$$\frac{1}{N} \sum_{m=1}^{N} \sum_{\alpha=1}^{l} \sum_{i=1}^{d_\alpha} y^*_{\alpha_i}(t) \left(\int_\Omega x^*_{\alpha_i}(ds) P^{(m)}(s, E) \right) = \sum_{\alpha=1}^{l} y_\alpha(t) x_\alpha(E),$$

or, by (4.52),

$$\frac{1}{N} \sum_{m=1}^{N} \sum_{\alpha=1}^{l} \sum_{i=1}^{d_\alpha} y^*_{\alpha_i}(t) x^*_{\alpha_{i+m}}(E) = \sum_{\alpha=1}^{l} y_\alpha(t) x_\alpha(E).$$

Using the fact that d_α is a divisor of N, we have

$$\sum_{\alpha=1}^{l} \left(\sum_{i=1}^{d_\alpha} y^*_{\alpha_i}(t) \right) \left(\frac{1}{d_\alpha} \sum_{i=1}^{d_\alpha} x^*_{\alpha_i}(E) \right) = \sum_{\alpha=1}^{l} y_\alpha(t) x_\alpha(E),$$

or, by (4.54),

$$\sum_{\alpha=1}^{l} \left(\sum_{i=1}^{d_\alpha} y^*_{\alpha_i}(t) \right) x_\alpha(E) = \sum_{\alpha=1}^{l} y_\alpha(t) x_\alpha(E);$$

and this implies (4.55) (put $E = \bar{E}_\alpha$).

In order to prove (4.53), we start from the trivial relation:

$$\int_\Omega P(t, ds) P_1^*(s, E) = \int_\Omega P_1^*(t, ds) P(s, E).$$

By (4.45) and (4.52), this becomes

$$\sum_{\alpha=1}^{l} \sum_{i=1}^{d_\alpha} \left(\int_\Omega P(t, ds) y^*_{\alpha_i}(s) \right) x^*_{\alpha_i}(E) = \sum_{\alpha=1}^{l} \sum_{i=1}^{d_\alpha} y^*_{\alpha_i}(t) \left(\int_\Omega x^*_{\alpha_i}(ds) P(s, E) \right)$$
$$= \sum_{\alpha=1}^{l} \sum_{i=1}^{d_\alpha} y^*_{\alpha_i}(t) x^*_{\alpha_{i+1}}(E),$$

and, putting $E = \bar{E}^*_{\alpha_{i+1}}$, we have

$$\int_\Omega P(t, ds) y^*_{\alpha_{i+1}}(s) = y^*_{\alpha_i}(t);$$

i.e., (4.53) is proved.

Thus we have proved (4.52), (4.53), (4.54) and (4.55). From (4.55), it is clear that each $\bar{E}^*_{\alpha_i}$ is contained in $\bar{E}_\alpha$, and (4.50) is a direct consequence of (4.53) (use Lemma 4.6).

Lastly, we shall prove (5.51). For any positive integers n and k we have, by (4.6) and (4.37),

$$P^{(nN+kd_\alpha)}(t, E) = \int_\Omega P^{(nN)}(t, ds) P^{(kd_\alpha)}(s, E) = \int_\Omega P_1^*(t, ds) P^{(kd_\alpha)}(s, E)$$
$$+ \int_\Omega S^{*(n)}(t, ds) P^{(kd_\alpha)}(s, E).$$

Since we have

$$\int_\Omega P_1^*(t, ds) P^{(kd_\alpha)}(s, E) = \int_\Omega x^*_{\alpha_i}(ds) P^{(kd_\alpha)}(s, E) = x^*_{\alpha_i}(E)$$

for any $t \in \bar{E}^*_{\alpha_i}$, and since, by (4.40),

$$\left| \int_\Omega S^{*(n)}(t, ds) P^{(kd_\alpha)}(s, E) \right| \leqq \underset{t \in \Omega, E \subset \Omega}{\text{l.u.b.}} \mid S^{*(n)}(t, E) \mid \leqq \frac{M}{(1+\epsilon)^n}$$

for $n = 1, 2, \cdots$, we have

$$\mid P^{(nN+kd_\alpha)}(t, E) - x^*_{\alpha_i}(E) \mid \leqq \frac{M}{(1+\epsilon)^n}$$

for $t \in \bar{E}^*_{\alpha_i}$, $k = 1, 2, \cdots, N/d_\alpha$; $n = 1, 2, \cdots$, where M and ϵ are positive constants which are independent of n and k. Hence we have (4.51), by a suitable change of M and ϵ.

Thus Theorem 11 is completely proved.

REMARK 3. We can also define *subergodic kernel* $E^*_{\alpha_i}$ in each subergodic part $\bar{E}^*_{\alpha_i}$; namely, $E^*_{\alpha_i}$ is the set of the smallest measure which satisfies $x^*_{\alpha_i}(E^*_{\alpha_i}) = 1$. If we suitably take E_α and $E^*_{\alpha_i}$ (which are all determined only up to a set of measure zero), then we have

COROLLARY.

$$E_\alpha = \sum_{i=1}^{d_\alpha} E^*_{\alpha_i}, \qquad E^*_{\alpha_i} = E_\alpha \bar{E}^*_{\alpha_i}, \tag{4.59}$$

$$P(t, E^*_{\alpha_{i+1}}) = 1. \qquad t \in E^*_{\alpha_i}. \tag{4.60}$$

PROOF. (4.59) is clear from (4.54), and (4.60) follows from (4.29), (4.50) and the second relation of (4.59).

It is to be noted that these E_α and $E^*_{\alpha_i}$ are exactly the final sets (ensembles finals) and their cyclic subsets which were discussed by W. Doeblin [1].

THEOREM 12.

$$\text{(4.61)} \qquad \underset{t \in \Omega}{\text{l.u.b.}}\ P^{(n)}(t, \Delta) \leqq \frac{M}{(1+\epsilon)^n}, \qquad n = 1, 2, \cdots,$$

where M and ϵ are positive constants which are independent of n.

PROOF. Since each $x^*_{\alpha_i}(E)$ satisfies

$$x^*_{\alpha_i}(\Delta) \leqq \sum_{i=1}^{d_\alpha} x^*_{\alpha_i}(\Delta) = d_\alpha \cdot x_\alpha(\Delta) = 0,$$

we have

$$P^*_1(t, \Delta) = \sum_{i=1}^{L} y^*_i(t) x^*_i(\Delta) \equiv 0$$

for any $t \in \Omega$, and consequently we have, by Theorem 10,

$$\underset{t \in \Omega}{\text{l.u.b.}}\ P^{(nN)}(t, \Delta) \leqq \frac{M}{(1+\epsilon)^n}, \qquad n = 1, 2, \cdots.$$

Since $P^{(n)}(t, \Delta)$ is monotone decreasing in n, we can deduce from this easily the relation (4.61) (by a suitable change of M and ϵ).

§4.7. Deduction of the condition (K) from the condition (D).

LEMMA 4.7. *Let us denote by $I(s_0)$ the closed interval $0 \leqq s \leqq s_0$. Then $Q(t, s) \equiv P(t, I(s))$ is Borel measurable as a function of two variables t and s in $0 \leqq t, s \leqq 1$.*

PROOF. By assumption, $Q(t, s)$ is Borel measurable in t if s is fixed, and if t is fixed $Q(t, s)$ is monotone increasing in s and is continuous on the right: $\lim_{s \to s_0+0} Q(t, s) = Q(t, s_0)$.

We shall prove that, for any α, the set $E(\alpha) = \underset{(t,s)}{E}\ [Q(t, s) < \alpha]$ is Borel measurable as a two-dimensional point set. For this purpose, put $E_s(\alpha) = \underset{t}{E}\ [Q(t, s) < \alpha]$. Since $Q(t, s)$ is monotone increasing in s, we have $E_{s_1}(\alpha) \supset E_{s_2}(\alpha)$ for $s_1 < s_2$, and consequently

$$E(\alpha) = \sum_{s \in \Omega} E_s(\alpha) \times I(s),$$

where $\times$ denotes the Cartesian product. Since $Q(t, s)$ is continuous in s on the right, the section of $E(\alpha)$ by the straight line $t = t_0$ is, if not empty, a semi-open interval of the form: $0 \leqq s < s_0$. Hence we have

$$E(\alpha) = \sum_{\substack{s = \text{rational} \\ s \in \Omega}} E_s(\alpha) \times I(s),$$

and this shows that $E(\alpha)$ is Borel measurable as a two-dimensional point set.

LEMMA 4.8.[19] *Let $K(t, E)$ and $n(t, E)$ be two kernels with bounded density:*

$$K(t, E) = \int_E k(t, s)\,ds, \qquad | k(t, s) | \leqq \| K \|,$$

$$N(t, E) = \int_E n(t, s)\,ds, \qquad | n(t, s) | \leqq \| N \|,$$

where $k(t, s)$ and $n(t, s)$ are both bounded measurable functions defined on $0 \leqq t, s \leqq 1$. If we consider the corresponding integral operators K and N, which map the Banach space $(\mathbf{M})$ into itself, then the integral operator NK defined by the kernel $\int_E \left(\int_0^1 k(t, u)n(u, s)\,du \right) ds$ is strongly completely continuous as an operator which maps $(\mathbf{M})$ into itself.

PROOF. For any $x(E) \in (\mathbf{M})$, put $y = K(x)$ and $z = N(y) = NK(x)$. The set function $z(E)$ is absolutely continuous:

$$z(E) = \int_E z'(s)\,ds,$$

and its density $z'(s)$ is given by

$$z'(s) = \int_0^1 y(du)n(u, s) = \int_0^1 \left(\int_0^1 x(dt)k(t, u) \right) n(u, s)\,du.$$

Hence $\| x \| \leqq 1$ implies

$$\underset{0 \leqq s \leqq 1}{\text{l.u.b.}} | z'(s) | \leqq \| K \| \cdot \| N \|$$

and

$$\int_{-\infty}^{+\infty} | z'(s + \delta) - z'(s) |\,ds \leqq \| K \| \int_{-\infty}^{+\infty} \left(\int_0^1 | n(u, s + \delta) - n(u, s) |\,du \right) ds$$

if we put $n(u, s) = 0$ for $s < 0$ and $s > 1$.

Consequently, we have

$$\lim_{\delta \to 0} \int_{-\infty}^{+\infty} | z'(s + \delta) - z'(s) |\,ds = 0$$

uniformly for all $x(E) \in (\mathbf{M})$ with $\| x \| \leqq 1$. Hence, by a theorem of A. Kolmogoroff [1] and M. Riesz [1], (if we consider $z'(s)$ as an element of the Banach space (L)) the totality of all $z'(s)$ corresponding to the unit sphere $\| x \| \leqq 1$ of $(\mathbf{M})$ is strongly compact in (L). In other words, NK is strongly completely continuous as an operator which maps $(\mathbf{M})$ into (L). Since (L) is isometric to a closed linear subspace of $(\mathbf{M})$, NK is also strongly completely continuous as an operator which maps $(\mathbf{M})$ into itself. Thus the proof of Lemma 4.8 is completed.

[19] K. Yosida, Y. Mimura and S. Kakutani [1].

Lemma 4.9. *If $P(t, E)$ satisfies the condition* (D), *then we have*

$$P^{(d)}(t, E) = \int_E q(t, s)\, ds + R(t, E), \tag{4.62}$$

where $q(t, s)$ is a bounded $\left(\leqq \frac{1}{\eta}\right)$ Borel measurable function defined for $0 \leqq t, s \leqq 1$, and

$$0 \leqq R(t, E) \leqq 1 - b \tag{4.63}$$

for any $t \in \Omega$ and $E \subset \Omega$.

Proof. Define $Q(t, s)$ newly by

$$Q(t, s) = P^{(d)}(t, I(s)).$$

By Lemma 4.7, $Q(t, s)$ is Borel measurable as a function of two variables t and s. Since $Q(t, s)$ is monotone in s for any t, $Q(t, s)$ is almost everywhere differentiable in s for any fixed t. If we put

$$p(t, s) = \lim_{n \to \infty} n\left(Q\left(t, s + \frac{1}{n}\right) - Q(t, s)\right)$$

and define $q(t, s)$ by

$$q(t, s) = \min\left(p(t, s), \frac{1}{\eta}\right),$$

then $q(t, s)$ is bounded $\left(\leqq \frac{1}{\eta}\right)$ and Borel measurable as a function of two variables t and s. We shall prove that the kernel $R(t, E)$ defined by

$$R(t, E) = P^{(d)}(t, E) - \int_E q(t, s)\, ds$$

satisfies (4.63) for any $t \in \Omega$ and $E \subset \Omega$.

For this purpose, we have only to notice that there exists for any $t \in \Omega$ a Borel set N_t of measure zero such that

$$P^{(d)}(t, E) = \int_E p(t, s)\, ds + P^{(d)}(t, N_t \cdot E)$$

for any Borel set $E \subset \Omega$. Then we have

$$\begin{aligned} 0 \leqq R(t, E) &= \int_E (p(t, s) - q(t, s))\, ds + P^{(d)}(t, N_t \cdot E) \\ &= \int_{E \cdot E_t(\eta)} p(t, s)\, ds + P^{(d)}(t, N_t \cdot E) \\ &\leqq P^{(d)}(t, E_t(\eta) + N_t), \end{aligned}$$

where $E_t(\eta) = \underset{s}{E}\left[p(t, s) > \frac{1}{\eta}\right]$. Since clearly mes $(E_t(\eta)) < \eta$ for any $t \in \Omega$, we have mes $(E_t(\eta) + N_t) < \eta$ and consequently, by (D), $0 \leqq R(t, E) \leqq 1 - b$ for any $t \in \Omega$ and $E \subset \Omega$.

THEOREM 13. *The condition* (D) *implies the condition* (K).

PROOF. By Lemma 4.9, we have

$$T^d = Q + R, \qquad \| R \| \leqq 1 - b, \tag{4.64}$$

where Q and R are the two integral operators which are defined by the kernels $Q(t, E) = \int_E q(t, s)\, ds$ and $R(t, E)$ respectively. If we expand $T^{dm} = (Q + R)^m$ in 2^m terms:

$$T^{dm} = Q^m + Q^{m-1}R + Q^{m-2}RQ + \cdots + QR^{m-1} + R^m,$$

then the terms which contain Q at least twice as factor are all strongly completely continuous. In order to see this, consider for example the term $RQRQ^{m-3}$. By (4.22), QR and Q^{m-3} are both integral operators with bounded density kernels. Hence, by Lemma 4.8, QRQ^{m-3} and consequently $RQRQ^{m-3}$ are strongly completely continuous. Since the number of terms which contain Q at most once as factor is $m + 1$, and since the norm of each such term is $\leqq (1 - b)^{m-1}$ by (4.64), we see that there exists, for each m, a strongly completely continuous operator V_m, which maps (**M**) into itself, such that

$$\| T^{dm} - V_m \| \leqq (m + 1)(1 - b)^{m-1}.$$

Since the right hand side converges to zero as $m \to \infty$, the proof of Theorem 13 is hereby completed.

REMARK. The converse of Theorem 13 is not true. To see this, take an arbitrary point $s_0 \in \Omega$, and define $P(t, E)$ by

$$\begin{aligned} P(t, E) &= 1 \quad \text{if} \quad s_0 \in E, \\ &= 0 \quad \text{if} \quad s_0 \bar{\in} E, \end{aligned}$$

for any $t \in \Omega$. This kernel $P(t, E)$ defines a strongly completely continuous integral operator, but the condition (*D*) is clearly not satisfied.

§4.8. Markoff's process with a finite number of possible states. Consider a Markoff's process with a finite number ($= N$) of possible states. Let p_{ij} ($i, j = 1, 2, \cdots, N$) be the transition probability that the i^{th} state is transferred to the j^{th} state after the elapse of a unit-time. Then the transition probability $p_{ij}^{(n)}$ that the i^{th} state is transferred to the j^{th} state after the elapse of n unit-times is given recurrently by

$$p_{ij}^{(n)} = \sum_{k=1}^{N} p_{ik} p_{kj}^{(n-1)}, \qquad p_{ij}^{(1)} = p_{ij}, \tag{4.65}$$

and we have always

$$p_{ij}^{(n)} \geqq 0, \qquad \sum_{j=1}^{N} p_{ij}^{(n)} = 1, \tag{4.66}$$

for $i, j = 1, 2, \cdots, N$; $n = 1, 2, \cdots$.

We shall investigate the asymptotic behavior of $p_{ij}^{(n)}$ for large n. This problem was discussed by many authors (see the introduction at the beginning of the paper), and sometimes direct methods were successful in this case. We shall, however, treat this problem as a special case of our general Markoff's process.

Consider the finitely valued function $p(t, s)$ defined in the square $0 \leqq t, s \leqq 1$ by

$$p(t, s) = N \cdot p_{ij} \quad \text{for} \quad \frac{i-1}{N} \leqq t < \frac{i}{N}, \quad \frac{j-1}{N} \leqq s < \frac{j}{N}, \tag{4.67}$$

$i, j = 1, 2, \cdots, N$ (in case $i = N$ or $j = N$, $<$ is to be replaced by $\leqq$). Then $P(t, E) = \int_E p(t, s)\, ds$ defines a simple Markoff's process on the interval $\Omega = (0, 1)$, and we have

$$P^{(n)}(t, E) = \int_E p^{(n)}(t, s)\, ds, \tag{4.68}$$

where

$$p^{(n)}(t, s) = N \cdot p_{ij}^{(n)} \quad \text{for} \quad \frac{i-1}{N} \leqq t < \frac{i}{N}, \quad \frac{j-1}{N} \leqq s < \frac{j}{N}, \tag{4.69}$$

$i, j = 1, 2, \cdots, N$ (in case $i = N$ or $j = N$, $<$ is again to be replaced by $\leqq$).

Thus the Markoff's process $P = (p_{ij})$ $(i, j = 1, 2, \cdots, N)$ is reduced to the continuous case $P = P(t, E)$. Since it is clear that the corresponding integral operator T is strongly completely continuous in this case, we have, by the results obtained above,

THEOREM 14. (i) *The limit*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} p_{ij}^{(m)} = p_{ij}^{(\infty)} \tag{4.70}$$

exists for any i and j, and there exists a constant M such that

$$\left| \frac{1}{n} \sum_{m=1}^{n} p_{ij}^{(m)} - p_{ij}^{(\infty)} \right| \leqq \frac{M}{n}, \qquad i, j = 1, 2, \cdots, N; n = 1, 2, \cdots. \tag{4.71}$$

(ii) *There exists a system of mutually disjoint ergodic parts $\bar{E}_\alpha$ $(\alpha = 1, 2, \cdots, l)$ such that*

$$\sum_{j \in \bar{E}_\alpha} p_{ij} = 1 \quad \textit{if} \quad i \in \bar{E}_\alpha, \tag{4.72}$$

(4.73) $$p_{ij}^{(\infty)} \text{ is independent of } i \text{ in each } \bar{E}_\alpha.$$

(iii) *Each ergodic part* $\bar{E}_\alpha$ *contains an ergodic kernel* E_α *such that*

(4.74) $$\sum_{j \in E_\alpha} p_{ij}^{(\infty)} = 1 \quad \text{if} \quad i \in \bar{E}_\alpha,$$

(4.75) $$\sum_{j \in E_\alpha} p_{ij} = 1 \quad \text{if} \quad i \in E_\alpha.$$

(iv) $\Delta = \Omega - \sum_{\alpha=1}^{l} E_\alpha$ *is called the dissipative part of* Ω *and we have*

(4.76) $$p_{ij}^{(n)} \leqq \frac{M}{(1+\epsilon)^n} \qquad j \in \Delta, i = 1, 2, \cdots, N; n = 1, 2, \cdots$$

with positive constants M *and* ϵ.

(v) *To each ergodic part* $\bar{E}_\alpha$ *there corresponds a positive integer* d_α *such that*

(4.77) $$| p_{ij}^{(nd_\alpha)} - d_\alpha p_{ij}^{(\infty)} | \leqq \frac{M}{(1+\epsilon)^n}, \quad i \in \bar{E}_\alpha, j = 1, 2, \cdots, N; n = 1, 2, \cdots$$

with positive constants M *and* ϵ. *Moreover, each* $\bar{E}_\alpha$ *contains* d_α *(mutually disjoint) subergodic parts* $\bar{E}^*_{\alpha_1}, \bar{E}^*_{\alpha_2}, \cdots, \bar{E}^*_{\alpha_{d_\alpha}}$ *such that*

(4.78) $$\sum_{j \in \bar{E}^*_{\alpha_{k+1}}} p_{ij} = 1 \quad \text{if} \quad i \in \bar{E}^*_{\alpha_k}, \qquad k = 1, 2, \cdots, d_\alpha \quad (\alpha_{d_\alpha+1} = \alpha_1).$$

(vi) *If we further put* $E^*_{\alpha_k} = E_\alpha \cdot \bar{E}^*_{\alpha_k}$ *for any* α *and* k, *then we have*

(4.79) $$\sum_{j \in E^*_{\alpha_{k+1}}} p_{ij} = 1 \quad \text{if} \quad i \in E^*_{\alpha_k}, \qquad k = 1, 2, \cdots, d_\alpha \quad (\alpha_{d_\alpha+1} = \alpha_1).$$

Added in proof. Recently, N. Dunford and B. J. Pettis [1] obtained some new results concerning weakly completely continuous operators defined on the space (L). Among others, they proved that if K and N are weakly completely continous operators which map (L) into itself, then NK is strongly completely continuous. This result is more precise than Lemma 4.8.

MATHEMATICAL INSTITUTE
OSAKA IMPERIAL UNIVERSITY.

BIBLIOGRAPHY

S. Banach [1]: *Théorie des opérations linéaires*, Warsaw, 1933.

G. Birkhoff [1]: *Dependent probabilities and the space (L)*, Proc. Nat. Acad. Sci., U. S. A., 24(1938), 154–159.

[2]: *The mean ergodic theorem*, Duke Math. Journ., 5(1939), 19–20.

W. Doeblin [1]: *Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples*, Bull. Soc. Math. Roumaine, 39–1(1937), 57–115; 39–2(1938), 3–61.

J. L. Doob [1]: *Stochastic processes with an integral valued parameter*, Trans. Amer. Math. Soc., 44(1938), 87–150.

N. Dunford [1]: *An ergodic theorem for n-parameter groups*, Proc. Nat. Acad. Sci., U. S. A., 25(1939), 195–196.

N. Dunford and B. J. Pettis [1]: *Linear operations on summable functions*, Trans. Amer. Math. Soc. 47(1940), 323–392.

M. Fréchet [1]: *Sur l'allure asymptotoque des densités itérées dans le problème des probabilités "en chaîne"*, Bull. Soc. Math. France, 62(1934), 68–83.

[2]: *Sur l'allure asymptotique de la suite des itérés d'un noyau de Frédholm.* Quart Journ. Math., 5(1934), 106–144.

[3]: *Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles.* Recherches théoriques modernes sur le calcul des probabilités, 2nd livre, Paris, 1938.

E. Hopf [1]: *Ergodentheorie, Ergebnisse der Math. und ihrer Grenzgebiete*, 5, Berlin, 1937.

B. Hostinsky [1]: *Méthodes générales du calcul des probabilités*, Mémorial des Sci. Math., 52, Paris, 1931.

S. Kakutani [1]: *Iteration of linear operations in complex Banach spaces*, Proc. Imp. Acad. Japan, 14(1938), 295–300.

[2]: *Mean ergodic theorem in abstract (L)-spaces*, Proc. Imp. Acad. Japan, 15(1939), 121–123.

[3]: *Weak topology and regularity of Banach spaces*, Proc. Imp. Acad. Japan, 15(1939), 167–173.

[4]: *Some results in the operator-theoretical treatment of the Markoff's process*, Proc. Imp. Acad. Japan, 15(1939), 260-264.

A. Kolmogoroff [1]: *Über die Kompaktheit der Funktionenmengen bei der Konvergenz im Mittel, Göttinger Nachrichten*, 1931, 60-63.

N. Kryloff and N. Bogolioùboff [1]: *Sur les propriétés en chaîne*, C.R. Paris, 204(1937), 1386-1388.

[2]: *Les propriétés ergodiques des suites des probabilités en chaîne*, C.R. Paris, 204(1937), 1454–1456.

D. Milman [1]: *On some criteria for the regularity of spaces of type (B)*, C.R. URSS, 20(1938), 243-246.

M. Nagumo [1]: *Einige analytische Untersuchungen in linearen normierten metrischen ringen*, Jap. Journ. of Math., 12(1936), 61–80.

J. V. Neumann [1]: *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sc., U. S. A., 18(1932), 587–642.

[2]: *Zur Operationsmethode in der klassischen Mechanik*, Annals of Math., 33(1932), 587–642.

B. J. Pettis [1]: *A proof that every uniformly convex space is reflexive*, Duke Math. Journ., 5(1939), 249–253.

F. Riesz [1]: *Über lineare Funktionalgleichungen*, Acta Math., 41(1918), 71–98.

[2]: *Some mean ergodic theorems*, Journ. of the London Math. Soc., 13(1938), 274–278.

M. Riesz [1]: *Sur les ensembles compacts de fonctions sommables*, Acta Szeged, 6(1932-34), 130–142.

C. Visser [1]: *On the iteration of linear operations in a Hilbert space*, Proc. Acad. Amsterdam, 41(1938), 487-495.

N. Wiener [1]: *Homogeneous chaos*, Amer. Journ. of Math., 61(1938), 71–98.

[2]: *The ergodic theorem*, Duke Math. Journ., 5(1939), 1–18.

K. Yosida [1]: *Abstract integral equation and the homogeneous stochastic process*, Proc. Imp. Acad. Japan, 14(1938), 286-291.

[2]: *Mean ergodic theorem in Banach spaces*, Proc. Imp. Acad. Japan, 14(1938), 292–294.

[3]: *Operator theoretical treatment of the Markoff's process*, Proc. Imp. Acad. Japan, 14(1938), 363–367. The second report under the same title, Proc. Imp. Acad. Japan, 15(1939), 127–130.

[4]: *Quasi-completely continuous linear functional equations*, Jap. Journ. of Math., 15(1939), 297–301.

K. Yosida and S. Kakutani [1]: *Application of mean ergodic theorem to the problem of Markoff's process*, Proc. Imp. Acad. Japan, 14(1938), 333–339.

K. Yosida, Y. Mimura and S. Kakutani [1]: *Integral operator with bounded kernel*, Proc. Imp. Acad. Japan, 14(1938), 359-362.

Annals of Mathematics
Vol. 42, No. 1, January, 1941

ON TYPES OF "WEAK" CONVERGENCE IN LINEAR NORMED SPACES

By Ingo Maddaus, Jr.

(July 28, 1939)

Introduction. In his paper "On a generalized notion of convergence in a Banach space," B. Vulich (Vulich 1, pp. 156 174) has pointed out that it is often possible to introduce a metric into a given concrete space with a given notion of convergence of a sequence of elements such that convergence in accordance with the metric is equivalent to the given convergence notion. However, it is not always possible to do this in terms of a metric which is a distance function (Vulich 1, p. 163). Consequently, Vulich has introduced a generalized metric by associating with each finite set or complex of a given linear set[1] a non-negative real number or norm. A set of six axioms imposed on this generalized metric and a limit notion defined in terms of it give rise to the concept of K-normed spaces.[2] The K-normed spaces are found to be special cases of Banach spaces (Banach, p. 53), and the K-convergence of a sequence of elements to a limiting element is found to imply Banach convergence to the same element. That is to say, K-convergence of a sequence to an element is stronger than Banach convergence of the sequence to the same element.

In this paper, also, the notion of the norm of a finite complex is employed. However, this norm is subjected to only three of the six axioms used in the definition of the K-normed spaces of Vulich. A limit notion for sequences is defined in terms of this norm. Among other things it is shown that point-wise convergence of a sequence of continuous functions to a continuous function on

[1] By a linear set will be meant the usual one given in Banach's book on page 26.

[2] A linear set X is said to be K-normed if there is associated with each finite set or complex of its elements a non-negative real number or norm, written $\|(x_1, x_2, \cdots, x_n)\|$, which satisfies the following axioms:

Axiom A. If $x_i = x_j$, then $\|(x_1, \cdots, x_i, \cdots, x_j, \cdots, x_n)\| = \|(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_j, \cdots, x_n)\| = \|(x_1, \cdots, x_i, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n)\|$.

Axiom B. $\|(x)\| = 0$ implies $x = \theta$, where θ is the null element of the linear set.

Axiom C. $\|(x_1, \cdots, x_n, x_1', \cdots, x_n')\| \leqq \|(x_1, \cdots, x_n)\| + \|(x_1', \cdots, x_n')\|$.

Axiom D. $\|(x_1, \cdots, x_n)\| \leqq \|(x_1, \cdots, x_n, x_{n+1})\|$.

Axiom E. $\|(x_1 + x_1', \cdots, x_n + x_n')\| \leqq \|(x_1, \cdots, x_n)\| + \|(x_1', \cdots, x_n')\|$.

Axiom F. $\|(ax_1, \cdots, ax_n)\| = |a| \cdot \|(x_1, \cdots, x_n)\|$.

By $x_n \to_k x$ it is meant that given any $e > 0$ there exists an $N(e)$ such that $n \geqq N(e)$ implies $\|(x_n - x, \cdots, x_{n+p} - x)\| < e$ for all $p \geqq 0$. The sequence $\{x_n\}$ is said to be K-convergent if given any $e > 0$ there exists an $N(e)$ such that if $n, m \geqq N(e)$ then $\|(x_n - x_m, \cdots, x_{n+p} - x_m)\| < e$ for all $p \geqq 0$.

A linear set is said to be a K-normed space if it is K-normed and if the limit notions are given by the definitions of the previous paragraph.

the finite closed interval reduces to the convergence of the sequence to the same function in terms of the norm of this paper. In all concrete cases considered it is seen that convergence of a sequence to an element in accordance with the limit notion to be defined is weaker than Banach convergence of the sequence to the element. Linear functionals and operations which are continuous in terms of this limit notion are studied.

1. Axioms and fundamental notions. Consider a linear set X of elements x such that there is associated with each finite subset or complex of elements of X a non-negative real number, called the norm of the complex and written $\| (x_1, x_2, \cdots, x_n) \|$ for the elements $x_1, x_2, \cdots, x_n$. This norm will satisfy the following axioms:

AXIOM A. If $x_i = x_j$, then $\| (x_1, \cdots, x_i, \cdots, x_j, \cdots, x_n) \| = \| (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_j, \cdots, x_n) \| = \| (x_1, \cdots, x_i, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) \|$.

AXIOM B. $\| (x) \| = 0$ implies $x = \theta$, where θ is the null element of the linear set X.

AXIOM 3. If $a_1, a_2, \cdots, a_n$ are real constants, then $\| (a_1x_1, a_2x_2, \cdots, a_nx_n) \| \leqq \max_{1 \leqq i \leqq n} | a_i | \cdot \| (x_1, x_2, \cdots, x_n) \|$.

A linear set with the norm of a finite complex defined and satisfying Axioms A, B, and 3 will be said to be H-normed.

DEFINITION 1. A sequence $\{x_n\}$ of elements of an H-normed set X will be said to be convergent in the H-sense, or to be H-convergent, if given any $e > 0$ and any infinite subsequence $\{x_{n_k}\}$ of $\{x_n\}$ there exists a $K_0 = K(e, \{x_{n_k}\})$ such that $k, s \geqq K_0$ imply

$$\| (x_{n_1} - x_{n_s}, \cdots, x_{n_k} - x_{n_s}) \| < e. \tag{1}$$

DEFINITION 2. A sequence $\{x_n\}$ of elements of an H-normed set X will be said to be convergent to the element $x \in X$ in the H-sense, or to have x as its H-limit, written $x_n \to_H x$ or $H\text{-}\lim_{n=\infty} x_n = x$, if given any $e > 0$ and any infinite subsequence $\{x_{n_k}\}$ of $\{x_n\}$ there exists a $K_0 = K(e, \{x_{n_k}\})$ such that $k \geqq K_0$ implies

$$\| (x_{n_1} - x, \cdots, x_{n_k} - x) \| < e. \tag{2}$$

An H-normed set for which the limit notions are given by Definitions 1 and 2 will be referred to as an H-normed space.

It is evident that in an H-normed space if a sequence converges or converges to an element x then every subsequence does likewise.

REMARK 1.1. It should be noted that our Axioms A and B are the Axioms A and B of K-normed spaces and Axiom 3 is a consequence of the axioms of K-normed spaces (Vulich 2, p. 61). Axioms A, B, and 3 are independent (Vulich 2, p. 60), and the examples which will be given in §2–5 will illustrate that there are H-normed spaces which are not K-normed according to the same norm.

The set of all real numbers forms an H-normed space if $\| (x_1, x_2, \cdots, x_n) \| = \inf (|x_1|, |x_2|, \cdots, |x_n|)$, where by "inf" is meant the smallest of the numbers involved. If in this same set $x_n \to_B x$ means that given any $e > 0$ there exists an $N(e)$ such that $n \geqq N(e)$ implies that $|x_n - x| < e$, then we have

THEOREM 1.1. *In the set of all real numbers $x_n \to_H x$ is equivalent to $x_n \to_B x$.*

PROOF. For any infinite subsequence $\{x_{n_k}\}$ of $\{x_n\}$, $\| (x_{n_1} - x, \cdots, x_{n_k} - x) \| = \inf (|x_{n_1} - x|, \cdots, |x_{n_k} - x|) \leqq |x_{n_k} - x|$. If $x_n \to_B x$ then this last expression approaches zero as $k \to \infty$. To prove the converse assume that $x_n \to_H x$ does not imply $x_n \to_B x$. Then there exists an infinite subsequence $\{x_{n_k}\}$ of $\{x_n\}$ and a positive number λ such that $|x_{n_k} - x| > \lambda > 0$ for $k = 1, 2, \cdots$. Then $0 < \lambda < \| (x_{n_1} - x, \cdots, x_{n_k} - x) \|$ for every integral k. This contradicts the assumption that $x_n \to_H x$.

REMARK 1.2. In general any linear set for which an F-metric (Banach, p. 35) is defined is an H-normed space if $\| (x_1, \cdots, x_n) \| = \inf [(x_1, \theta), \cdots, (x_n, \theta)]$. By Theorem 1.1 $x_n \to_H x$ is equivalent to $x_n \to x$ in accordance with the F-metric. Hence, any linear normed space[3] is also an H-normed space, and when the norm of a finite complex is defined as in this paragraph norm convergence of a sequence to an element is equivalent to H-convergence to the same element. Under the same circumstances norm convergence of a sequence is equivalent to H-convergence of the sequence.

The following theorems relate to H-normed spaces.

THEOREM 1.2. *If $x_n \to_H x$ and $x_n = x'_n$ for every n, then $x'_n \to_H x$.*

THEOREM 1.3. *If $x_n \to_H x$, then for any $x' \in X$, $x_n + x' \to_H x + x'$.*

REMARK 1.3. In Remark 3.1 an example will be given which will show that $x_n \to_H x$ and $y_n \to_H y$ do not imply that $x_n + y_n \to_H x + y$.

Because of Theorem 1.3 it is evident that the topology of an H-normed space is a uniform topology.

THEOREM 1.4. $\| (ax_1, \cdots, ax_n) \| = |a| \cdot \| (x_1, \cdots, x_n) \|$.

PROOF. $\| (ax_1, \cdots, ax_n) \| \leqq |a| \cdot \| (x_1, \cdots, x_n) \| = |a| \cdot \| (aa^{-1}x_1, \cdots, aa^{-1}x_n) \| \leqq |a| \cdot |a^{-1}| \cdot \| (ax_1, \cdots, ax_n) \| = \| (ax_1, \cdots, ax_n) \|$. This completes the proof.

THEOREM 1.5. *If $\{a_n\}$ is a sequence of real numbers such that $|a_n| \leqq M$, where M is independent of n, and if $x_n \to_H \theta$, then $a_n x_n \to_H \theta$.*

PROOF. Choose any infinite subsequence $\{a_{n_k} x_{n_k}\}$ of $\{a_n x_n\}$. Then $\| (a_{n_1} x_{n_1}, \cdots, a_{n_k} x_{n_k}) \| \leqq M \| (x_{n_1}, \cdots, x_{n_k}) \|$. Since $x_n \to_H \theta$ it follows that the right hand side of this inequality approaches zero as $k \to \infty$.

COROLLARY 1.51. *If $\{a_n\}$ is a sequence of real numbers such that $a_n \to_B 0$ and if $x_n \to_H \theta$, then $a_n x_n \to_H \theta$.*

COROLLARY 1.52. *If a is a real number and $x_n \to_H x$, then $ax_n \to_H ax$.*

[3] A linear normed space is a linear set X with the property that there is associated with each element x a non-negative real number or norm, designated by $\| x \|$, which is such that (1) $\| x \| = 0$ implies $x = \theta$, (2) $\| x + y \| \leqq \| x \| + \| y \|$, (3) $\| tx \| = |t| \cdot \| x \|$ for each real t. A sequence $\{x_n\}$ of X is said to be convergent to $x \in X$ in the norm or Banach sense if $\| x_n - x \| \to 0$ as $n \to \infty$.

2. The space (CH). Consider the set of all continuous functions defined on the closed interval $(0, 1)$, and let the norm of a finite complex of elements be given by $\| (x_1, \cdots, x_n) \| = \max_{0 \leqq t \leqq 1} \inf (| x_1(t) |, \cdots, | x_n(t) |)$. With this norm the set becomes an H-normed space and it will be designated by (CH).

THEOREM 2.1. *In the space* (CH) *a necessary and sufficient condition that* $x_n \rightarrow_H x$ *is that* $x_n(t) \rightarrow x(t)$ *for each* $t \in (0, 1)$.

PROOF. Because of Theorem 1.3 it suffices to prove this theorem for the case $x_n \rightarrow_H \theta$.

Necessity. Assume on the contrary that for some t_0, $\overline{\lim} | x_n(t_0) | = 2\lambda > 0$. Then there is a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ with the property that $| x_{n_k}(t_0) | > \lambda$ for each integral k. Therefore, $\| (x_{n_1}, \cdots, x_{n_k}) \| \geqq \inf (| x_{n_1}(t_0) |, \cdots, | x_{n_k}(t_0) |) > \lambda > 0$ for each integral k. Consequently, θ is not an H-limit of $\{x_n\}$, in contradiction to the assumption of the theorem.

Sufficiency. Assume that there is a subsequence of $\{x_n\}$, which will also be designated by $\{x_n\}$, such that $\| (x_1, \cdots, x_n) \|$ is greater than a positive number λ for $n = 1, 2, \cdots$. Designate $\inf (| x_1(t) |, \cdots, | x_n(t) |)$ by $f_n(t)$. For a fixed value of n it is easily shown that $f_n(t)$ is a continuous function on the closed interval $(0, 1)$. Hence, there exists a t_n such that $f_n(t_n) = \max_{0 \leqq t \leqq 1} | f_n(t) | = \| (x_1, \cdots, x_n) \|$. By the Weierstrass–Bolzano Theorem it is possible to choose a subsequence $\{t_{n_k}\}$ of $\{t_n\}$ and a point t_0 such that $\lim_{k \to \infty} t_{n_k} = t_0$.

Each $f_n(t_0) \geqq \lambda$. For if $f_{n'}(t_0) < \lambda$ there would exist an interval I about t_0 such that if $t \in I$ then $f_{n'}(t) < \lambda$. Now choose $n_k > n'$ and such that $t_{n_k} \in I$; then $f_{n_k}(t_{n_k}) < f_{n'}(t_{n_k}) < \lambda$, which is a contradiction. But $f_n(t) \leqq | x_n(t) |$ for all n and t, so $| x_n(t_0) | > f_n(t_0) \geqq \lambda$, in contradiction to the assumption that $x_n(t) \rightarrow 0$ for all $t \in (0, 1)$.

REMARK 2.1. If $x_1, x_2, \cdots, x_n$ and $x_1', x_2', \cdots, x_n'$ are two finite complexes of an H-normed space it does not follow that $\| (x_1 + x_1', x_2 + x_2', \cdots, x_n + x_n') \| \leqq \| (x_1, x_2, \cdots, x_n) \| + \| (x_1', x_2', \cdots, x_n') \|$. For consider the following complexes from the H-normed space (CH): $x = 2, x = 5$ and $x' = 5$, $x' = 2$. In this case the inequality sign is actually reversed. If the sign of x_1 is changed then the inequality sign is again reversed. Therefore, unlike the norm of the K- normed spaces of B. Vulich, the norm of a finite complex of a H-normed space does not satisfy a "triangle property," nor does it satisfy a "triangle property" with the sign reversed.

REMARK 2.2. In the space (CH) it can readily be shown by methods similar to those employed in the proof of Theorem 2.1 that a necessary and sufficient condition for the convergence of a sequence in the H-sense is: $| x_n(t) - x_m(t) | \rightarrow 0$ for each $t \in (0, 1)$. There are numerous familiar examples from the space (CH) which illustrate that H-convergence of a sequence does not imply that the sequence has an H-limit which is of the space. That is to say, an H-normed space is not necessarily complete with respect to the H-convergence notions.

REMARK 2.3. The norm of a finite number of elements of an H-normed space

is not necessarily a continuous function of its arguments; i.e., $x_n^{(i)} \to_H x^{(i)}$ for $i = 1, 2, \cdots, p$ does not imply $\lim_{n=\infty} \| (x_n^{(1)}, \cdots, x_n^{(p)}) \| = \| (x^{(1)}, \cdots, x^{(p)}) \|$. For consider the following sequence from (CH):

$$x_n(t) = x_n(0) = 0 \text{ for } 2/n \leqq t \leqq 1,$$
$$x_n(1/n) = n,$$

and $x_n(t)$ linear elsewhere ($n = 1, 2, \cdots$) (Caratheodory, p. 171). This sequence is such that $x_n \to_H \theta$. Let a second sequence be given by $x_n'(t) = 1$ for each n and for $t \in (0, 1)$. Then $x_n' \to_H 1$. Moreover, $\| (x_n, x_n') \| = 1$ for each n, whereas $\| (\theta, 1) \| = 0$.

3. The space (L^pH), $p \geqq 1$. The norm of a finite complex which was used in the previous paragraph to make the set of all continuous functions on the finite closed interval an H-normed space was obtained by operating on the greatest lower bound of the absolute values of the elements involved with the norm which is generally employed to make the set of all continuous functions a Banach space (Banach, p. 11). To define $\| (x_1, \cdots, x_n) \|$ in an analogous manner for the set of all functions defined on the closed interval (0, 1) and whose pth powers are summable set it equal to $\left\{ \int_0^1 [\inf (| x_1(t) |, \cdots, | x_n(t) |)]^p \, dt \right\}^{1/p}$, where by "inf" is meant the greatest lower bound of the functions involved at every point of (0, 1).[4] With this interpretation of the norm of a finite complex the set of functions considered is an H-normed space. This space will be referred to as the space (L^pH), $p \geqq 1$. It will be seen from what follows that the H-convergence of a sequence of elements of (L^pH), $p \geqq 1$, to an element of the space is extremely weak.

LEMMA 3.1. *If $\{f_n\}$ is a sequence of summable functions defined on* (0, 1) *and such that*

(1) $f_n(t) \geqq 0$ *for each* n,

(2) $\{f_n(t)\}$ *is non-increasing,*

(3) $\lim_{n=\infty} \int_0^1 f_n(t) \, dt = 0$,

then $\lim_{n=\infty} f_n(t) = 0$ *except on a set of zero measure.*

THEOREM 3.1. *In* (L^pH), $p \geqq 1$, $x_n \to_H x$ *is equivalent to either of the following statements*:

(1) *For any subsequence* $\{x_{n_k}\}$ *of* $\{x_n\}$, $\inf_k | x_{n_k}(t) - x(t) | = 0$ *except on a set of measure zero which depends on the subsequence.*

---

[4] All the work of this section would go through without difficulty if the "inf" should be defined as the greatest lower bound of the functions involved except on a set of zero measure.

(2) *For any subsequence* $\{x_{n_k}\}$ *of* $\{x_n\}$ *there is a set* E *of zero measure such that for* $t_0 \notin E$ *there is a subsequence* $\{x_{n_{k_l}}\}$ *of* $\{x_{n_k}\}$ *such that* $\lim_{l\to\infty} x_{n_{k_l}}(t_0) = x(t_0)$.

PROOF. Because of Theorem 1.3 it is sufficient to prove this theorem for the case $x_n \to_H \theta$. We first show that $x_n \to_H \theta$ implies condition (2). Let $f_n^s(t) = [\inf(|x_s(t)|, \cdots, |x_n(t)|)]^p$ for each $t \in (0, 1)$. If s is constant then by Lemma 3.1 $\lim_{n=\infty} f_n^s(t) = 0$ except on a set E_s of zero measure. Let $E = \sum_{s=1}^{\infty} E_s$; then E is of zero measure and it is the set mentioned in the statement of the theorem. For consider any t_0 of the complement of E; $\lim_{n=\infty} f_n^s(t_0) = 0$ for each s, and in particular for $s_0 = 1$. Hence, there exists an integer $s_1 > s_0$ such that $|x_{s_1}(t_0)| < 1$. Consider $\{f_n^{s_1+1}\}$; $\lim_{n=\infty} f_n^{s_1+1}(t_0) = 0$, and so there exists an integer $s_2 > s_1$ such that $|x_{s_2}(t_0)| < \frac{1}{2}$. In this way an increasing sequence $\{s_i\}$ of integers can be built up in such a way that $|x_{s_i}(t_0)| < 1/i$. Then $\lim_{i=\infty} x_{s_i}(t_0) = 0$. Therefore, E is the set mentioned in statement (2). Since $x_{n_k} \to_H \theta$ is a consequence of $x_n \to_H \theta$, and since we may apply the above argument to $\{x_{n_k}\}$, our contention is proved.

It is obvious that statement (2) implies statement (1).

Now suppose that (1) holds when $x = \theta$. Let $\{x_{n_k}\}$ be a subsequence of $\{x_n\}$ and let $f_k(t) = [\inf(|x_{n_1}(t)|, \cdots, |x_{n_k}(t)|)]^p$. Then (1) states that $\inf_k |x_{n_k}(t)| = \lim_{k=\infty} [f_k(t)]^{1/p} = 0$ except on a set E of zero measure. Hence, $\lim_{k=\infty} f_k(t) = 0$ except on E and $\lim_{k=\infty} \int_0^1 f_k(t)\, dt = 0$. Therefore, $x_n \to_H \theta$.

REMARK 3.1. The following example from (L^pH) shows that in an H-normed space the H-limit of a sequence is not necessarily unique. Let $\{x_n\}$ be defined by $x_n(t) = \frac{1}{2}(1 + r_n(t))$, where $r_n(t) = \operatorname{sgn}(\cos 2\pi nt)$.[5] If $E_n = E(x_n(t) = 1)$, then $\{E_n\}$ is a sequence of sets of closed intervals with no points common to any pair. This sequence has the property that the measure of the points common to each E_{n_k} of any subsequence $\{E_{n_k}\}$ is zero. Consequently, each $\{x_{n_k}\}$ of $\{x_n\}$ has the property that $\lim_{k=\infty} \int_0^1 [\inf(|x_{n_1}(t)|, \cdots, |x_{n_k}(t)|)]^p\, dt = 0$, and so $x_n \to_H \theta$. Now consider the sequence $\{x_n - 1\}$. By a study of the complements of the E_n sets—the sets on which $x_n - 1 = -1$—it is readily seen that $x_n - 1 \to_H \theta$. By Theorem 1.3 it follows that $x_n \to_H 1$. Hence the sequence $\{x_n\}$ has the H-limits θ and 1.

Consider the sequence $\{x_n\}$ just defined together with the sequence $\{x'_n\}$, where $x'_n = 1 - x_n$. Since $\{x'_n\}$ has the H-limits θ and 1 it follows that $H\text{-}\lim_{n=\infty} x_n + H\text{-}\lim_{n=\infty} x'_n$ may equal either θ or 1. But $H\text{-}\lim_{n=\infty} (x_n + x'_n) = 1$, so in an H-normed space $x_n \to_H x$ and $y_n \to_H y$ do not imply $(x_n + y_n) \to_H x + y$.

[5] The r_n functions are closely related to the well known Rademacher functions.

REMARK. 3.2. By giving different interpretations to $\| (x_1, \cdots, x_n) \|$ in the set of all functions defined on (0, 1) and whose pth powers are summable the H-limit notion will be found to have different meanings. For example, in the case of all summable functions defined on (0, 1) let $\| (x_1, \cdots, x_n) \| = \max_{0 \leq t \leq 1} \inf \left(\left| \int_0^t x_1(t)\, dt \right|, \cdots, \left| \int_0^t x_n(t)\, dt \right| \right)$. The set of all summable functions on (0, 1) is an H-normed space according to this norm and it will be designated by $(LH)_1$. Since $\int_0^t x(t)\, dt$ is a continuous function of t on (0, 1) it is evident from Theorem 2.1 that $x_n \to_H x$ is now equivalent to $\lim_{n\to\infty} \int_0^t x_n(t)\, dt = \int_0^t x(t)\, dt$ for every $t \in (0, 1)$.

4. The space (cH). Consider the set of all convergent sequences $\{\xi_i\}$ of real numbers. Let an element x_n of the set be designated by $\{\xi_{ni}\}$ and let $\| (x_1, \cdots, x_n) \| = \sup_{1 \leq i \leq \infty} | a_i^n |$, where by "sup" is meant the least upper bound of the numbers involved, and $a_i^n = \inf (| \xi_{1i} |, \cdots, | \xi_{ni} |)$. Then the set is an H-normed space and it will be designated by (cH).

THEOREM 4.1. *A necessary and sufficient condition that* $x_n \to_H x$ in (cH) *is that*

$$\text{(a)} \qquad \lim_{n\to\infty} \xi_{ni} = \xi_i \quad \text{for} \quad i = 1, 2, \cdots,$$

$$\text{(b)} \qquad \lim_{n\to\infty} \lim_{i\to\infty} \xi_{ni} = \lim_{i\to\infty} \xi_i .$$

PROOF. This theorem can be derived from Theorem 2.1 since (cH) can be put into a 1-1 norm preserving correspondence with a subset of (CH). The correspondence is $\{\xi_i\}$ to $f(t)$ where $f(1/i) = \xi_i$, $f(0) = \lim_{i\to\infty} \xi_i$ and $f(t)$ is linear elsewhere.[6]

5. The space $(l^p H)$, $p \geqq 1$. Consider the set of all sequences $\{\xi_i\}$ of real numbers such that $\sum_{i=1}^{\infty} | \xi_i |^p < \infty$. Let any element x_n be designated by $\{\xi_{ni}\}$, and let $\| (x_1, \cdots, x_n) \| = [\sum_{i=1}^{\infty} | a_i^n |^p]^{1/p}$, where $a_i^n = \inf (| \xi_{1i} |, \cdots, | \xi_{ni} |)$. Then the set is an H-normed space and it will be designated by $(l^p H)$, $p \geqq 1$.

THEOREM 5.1. *A necessary and sufficient condition that* $x_n \to_H x$ *in* $(l^p H)$ *is that* $\lim_{n\to\infty} \xi_{ni} = \xi_i$ *for* $i = 1, 2, \cdots$.

PROOF. This theorem can be derived from statement (1) of Theorem 3.1 since $(l^p H)$, $p \geqq 1$, can be put into a 1-1 norm preserving correspondence with a subset of $(L^p H)$, $p \geqq 1$. The correspondence is $\{\xi_i\}$ to $f(t)$, where $f(t) = i(i + 1)\xi_i$ for $1/(1 + i) < t < 1/i$ and $f(t) = 0$ elsewhere.

[6] The proofs of Theorems 4.1 and 5.1 were suggested by the referee. They are used here because they are much simpler than those originally given by the author.

6. Weak sequential convergence. In the last four sections certain concrete sets of elements have been considered as H-normed spaces by giving definite meanings to the norms of finite subsets of elements. Each of these sets may also be considered as a Banach space by proper definition of the norm for single elements. Banach has given necessary and sufficient conditions for weak sequential convergence[7] in each case considered (Banach, pp. 134–137). An examination of our Theorems 2.1, 3.1, 4.1, and 5.1 will readily show that the necessary and sufficient conditions for $x_n \rightarrow_H x$ are in each case weaker than the necessary and sufficient conditions for weak sequential convergence. That is to say, $x_n \rightarrow_H x$ for the sets considered has been made weaker than weak sequential convergence by the interpretations given the norms of finite complexes. In this section different interpretations of the norms of finite complexes on the same sets considered in the previous sections will be given, and the resulting H-convergence notions will be found to be equivalent to weak sequential convergence.

To this end it is advisable to prove first a general theorem on Banach spaces and then to give appropriate interpretations to the norms of finite complexes in the different cases to be considered. Let E be a Banach space and let $\bar{E}$ be its conjugate, or the space of all linear limited functionals defined on E (Banach, p. 54). If $\{f_n\}$ is a sequence of and f is a single linear limited functional defined on E, then $f_n \rightarrow f$ will always be taken to mean $f_n(x) \rightarrow f(x)$ for every $x \in E$; i.e., $f_n \rightarrow f$ will mean that $\{f_n\}$ converges weakly to f (Banach, p. 122). Bounded sets of $\bar{E}$ will be weakly compact if every sequence chosen from a bounded (in the norm sense) set of linear limited functionals defined on E is such that there is a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ and another linear limited functional f_0 defined on E such that $f_{n_k} \rightarrow f_0$. It should be noted that if $f_n \rightarrow f$ and $\|f\| \leqq M$, where M is independent of n, then $\|f\| \leqq \underline{\lim} \|f_n\| \leqq M$ (Banach, p. 123).

In the Banach space E let $\|(x_1, \cdots, x_n)\| = \max_{\|f\|=1} G_n(f)$, where f is a linear limited functional on E and where $G_n(f) = \inf(|f(x_1)|, \cdots, |f(x_n)|)$. $G_n(f)$ is a real valued function defined on $\bar{E}$. Because of the introduction of this norm for a finite complex the elements which form the Banach space E also form an H-normed space (EH). In this section the notions of H-convergence will be in terms of this norm.

Lemma 6.1. *If $f_s \rightarrow f$ then $G_n(f_s) \rightarrow G_n(f)$.*

Proof. This lemma follows readily from the fact that each $G_n(f)$ involves only a finite number of elements of E and from the definition of $f_s \rightarrow f$.

Let F be the subset of $\bar{E}$ such that if $f \in F$ then $\|f\| = 1$. Let F_0 be the closed extension of F in the sense that if $f_n \in F$ and $f_n \rightarrow f$ then $f \subset F_0$.

Lemma 6.2. *If the Banach space E is such that in $\bar{E}$ bounded sets are weakly compact, then $\max_{\|f\|=1} G_n(f)$ is attained by some element $f_0 \subset F_0$.*

[7] The sequence $\{x_n\}$ of a space X is said to be convergent to an element $x \in X$ in the weak sequential manner if for every linear (homogeneous and additive) and continuous functional $f(x)$ defined on X, $\lim f(x_n) = f(x)$ (Banach, p. 133).

PROOF. Consider any $G_{n_0}(f)$ and a sequence $\{f_s\}$ of elements of F, and let $\lim_{s\to\infty} G_{n_0}(f_s) = \max_{||f||=1} G_{n_0}(f)$. Each $G_{n_0}(f_s)$ is actually equal to the absolute value of f_s at some point of E; for by definition each $G_{n_0}(f_s)$ involves only a finite number of elements of E. Therefore, $G_{n_0}(f_s) = |f_s(x_i)|$, where i assumes one of the values $1, 2, \cdots, n_0$. Since the sequence $\{f_s\}$ is infinite there is some one x_{i_0} and an infinite subsequence $\{f_{s_p}\}$ of $\{f_s\}$ such that $G_{n_0}(f_{s_p}) = |f_{s_p}(x_{i_0})|$ for each integral p. Since each $||f_{s_p}|| = 1$ there is a subsequence $\{f_{s_{p_j}}\}$ of $\{f_{s_p}\}$ and an f_0, of unit norm or less, such that $f_{s_{p_j}} \to f_0$. By Lemma 6.1 it follows that $\lim_{j\to\infty} G_{n_0}(f_{s_{p_j}}) = G_{n_0}(f_0)$. This completes the proof.

THEOREM 6.1. *If E is a Banach space such that in $\bar{E}$ bounded sets are weakly compact then $x_n \to_H x$ is equivalent to the weak sequential convergence of $\{x_n\}$ to x.*

PROOF. This theorem is a consequence of Lemmas 6.1 and 6.2 and its proof follows almost word by word the proof of Theorem 2.1.

It is well-known that the sets of elements which form the spaces (CH), (L^pH), (l^pH), $(p \geqq 1)$, and (cH) are also elements of separable Banach spaces if the norm for single elements is properly chosen. Banach states (Banach, p. 123) that each separable Banach space E has a conjugate space whose bounded sets are weakly compact as sets of functionals over E. Consequently, if the norm of a finite complex is defined as in this section for any one of the concrete Banach spaces just mentioned, then $x_n \to_H x$ is equivalent to the weak sequential convergence of $\{x_n\}$ to x.

7. Comparison of limit notions. It is of interest to note how the gap between the H-convergence in H-normed spaces and the K-convergence in K-normed spaces may be bridged. It is first necessary to introduce a new limit notion.

DEFINITION 1^0. A sequence $\{x_n\}$ of elements of an H-normed set X will be said to be convergent in the G-sense, or to be G-convergent, if given any $e > 0$ and any infinite subsequence $\{x_{n_k}\}$ of $\{x_n\}$ there exists a $K_0 = K(e, \{x_{n_k}\})$ such that for $k \geqq K_0$ there exists a $P_0 = P(e, k, \{x_{n_k}\})$ such that $k \geqq K_0$, $s \geqq K_0$ and $p \geqq P_0$ imply

$$(1^0) \quad ||(x_{n_k} - x_{n_s}, \cdots, x_{n_{k+p}} - x_{n_s})|| < e.$$

DEFINITION 2^0. A sequence $\{x_n\}$ of elements of an H-normed set X will be said to be convergent to the element $x \in X$ in the G-sense, or to have x as its G-limit, written $x_n \to_G x$ or $G\text{-}\lim_{n\to\infty} x_n = x$, if given any $e > 0$ and any infinite subsequence $\{x_{n_k}\}$ of $\{x_n\}$ there exists a $K_0 = K(e, \{x_{n_k}\})$ such that for $k \geqq K_0$ there exists a $P_0 = P(e, k, \{x_{n_k}\})$ such that $k \geqq K_0$ and $p \geqq P_0$ imply

$$(2^0) \quad ||(x_{n_k} - x, \cdots, x_{n_{k+p}} - x)|| < e.$$

THEOREM 7.1. *In a K-normed set $x_n \to_K x$ is equivalent to $x_n \to_G x$.*

PROOF. If $x_n \to_G x$ then given any $e > 0$ there exists an $N(e)$ and if $n \geqq N(e)$ there exists a $P(e, n)$ such that $n \geqq N(e)$ and $p \geqq P(e, n)$ imply

$\| (x_n - x, \cdots, x_{n+p} - x) \| < e$. But for all $0 \leqq i \leqq p$, $\| (x_n - x, \cdots, x_{n+i} - x) \| \leqq \| (x_n - x, \cdots, x_{n+p} - x) \| < e$, because of Axiom D of K-normed sets. Hence, the required inequality holds for every $i \geqq 0$, and so $x_n \to_K x$.

To prove the converse note that $x_n \to_K x$ implies $x_{n_i} \to_K x$ for every subsequence $\{x_{n_i}\}$ of $\{x_n\}$. From this fact it is immediately evident that $x_n \to_K x$ implies $x_n \to_G x$.

THEOREM 7.2. *If in a K-normed set a sequence $\{x_n\}$ is K-convergent then it is G-convergent, and conversely.*

In the proofs of Theorems 2.1, 3.1, 4.1, 5.1, and 6.1 we made use of a property which when generalized to cover all cases may be written as

PROPERTY 1. $\| (x_1, \cdots, x_n, x_{n+1}) \| \leqq \| (x_1, \cdots, x_n) \|$.

This property is actually Axiom D of K-normed sets with the inequality sign reversed. Since the norms of the spaces (CH), (L^pH), (l^pH), $(p \geqq 1)$, (cH) and (EH) satisfy Property 1 it is evident that they are not K-normed spaces.

THEOREM 7.3. *In an H-normed space whose norm satisfies Property 1 $x_n \to_H x$ is equivalent to $x_n \to_G x$.*

PROOF. The proof of this theorem depends on Property 1 in much the same manner that the proof of Theorem depends on Axiom D.

THEOREM 7.4. *In an H-normed space whose norm satisfies Property 1 if a sequence is G-convergent then it is H-convergent, and conversely.*

REMARK 7.1. Consequently, the G-convergence notions may be used in place of the H-convergence notions and the K-convergence notions in the mutually exclusive classes of H-normed sets satisfying Property 1 and of K-normed sets, respectively.

8. H^*-normed spaces. By an H^*-normed space will be meant an H- normed space whose elements also form a linear normed space and in which convergence of a sequence $\{x_n\}$ to x in the norm sense (Banach convergence), i.e., $\| x_n - x \| \to 0$—written $x_n \to_B x$—implies $x_n \to_H x$. If an H-normed space satisfies Property 1 and if it is linear normed with respect to the norm of a finite complex when that norm operates on single elements then it is evidently an H^*-normed space. The K-normed spaces are not necessarily H-normed spaces (Vulich 1, p. 163) for the K-limit notion is in general stronger than the limit notion in the norm sense.

REMARK 8.1. The example of Remark 3.2 is an H-normed space satisfying Property 1. Moreover, the set of elements of that example form a linear normed space when the norm $\| x \| = \max_{0 \leqq t \leqq 1} \left| \int_0^t x(u)\, du \right|$ is used, and so it is an H^*-normed space. However, this is not the usual norm employed to make the set of summable functions on (0, 1) a linear normed space—the usual norm is $\| x \| = \int_0^1 | x(u) |\, du$. Nevertheless, it is evident that the space $(LH)_1$ is an H^*-normed space when $x_n \to_B x$ is in accordance with the second norm just given and when $x_n \to_H x$ is with respect to the norm of a finite complex of $(LH)_1$.

By an $\bar{H}$-normed space will be meant an H^*-normed space in which the H-limit is unique and satisfies the condition: $x_n \to_H x$ and $y_n \to_H y$ imply $(x_n + y_n) \to_H (x + y)$.

REMARK 8.2. The space (L^pH), $p \geqq 1$, is an H^*-normed space which is not $\bar{H}$-normed.

LEMMA 8.1. *If $\{a_n\}$ is a sequence of real numbers such that $a_n \to_B a$ and if x is an element of an H-normed space whose norm satisfies Property 1, then $a_n x \to_H ax$.*

THEOREM 8.1. *If $\{x_n\}$ is a sequence of elements of an $\bar{H}$-normed space X whose norm satisfies Property 1, if $x_n \to_H x$ where $x \in X$, and if $\{a_n\}$ is a sequence of real numbers such that $a_n \to_B a$, then $a_n x_n \to_H ax$.*

PROOF. $a_n x_n - ax = (a_n x_n - a_n x) + (a_n x - ax)$. From Theorem 1.5 it follows that the first expression on the right approaches θ in the H-sense, and from Lemma 8.1 it follows that the second expression does likewise. The theorem follows from the fact that the space is $\bar{H}$-normed.

REMARK 8.3. An $\bar{H}$-normed space whose norm satisfies Property 1 does not have the property that the derived class is closed; that is to say, if $x_n^m \to_H x_n$ for $n = 1, 2, \cdots$, and if $x_n \to_H x$, it does not follow that there exists a subsequence $\{x_n^{m_n}\}$, where $m_1 < m_2 < \cdots < m_n < \cdots$, such that $x_n^{m_n} \to_H x$. For consider the following example from the space of all continuous functions defined on (0, 1) with the norm defined as in §6. It has been shown in §6 that in this space $x_n \to_H x$ is equivalent to the weak sequential convergence of $\{x_n\}$ to x; i.e., $x_n \to_H x$ is equivalent to $x_n(t) \to x(t)$ for each $t \in (0, 1)$ and to the existence of a constant $M > 0$ such that $\max_{0 \leqq t \leqq 1} | x_n(t) | \leqq M$ (Banach, p. 134). Let $\{x_n\}$ be defined as follows:

$$x_n(t) = x_n(0) = 0 \text{ for } 2/n \leqq t \leqq 1,$$
$$x_n(1/n) = 1,$$

$x_n(t)$ linear elsewhere. This sequence obviously has θ as its weak sequential limit. Let x_n^m be such that

$$x_n^m(t) = x_n(t) \text{ for } 0 \leqq t \leqq 2/n,$$
$$x_n^m(t) = 0 \text{ for } 2/n + 2/m \leqq t \leqq 1,$$
$$x_n^m(2/n + 1/m) = n,$$

$x_n^m(t)$ linear elsewhere. For each fixed value of n, $x_n^m \to x_n$ in the weak sequential sense. But since weak sequential convergence implies boundedness of the norms of elements it follows that it is impossible to find a sequence of integers $m_1 < m_2 < \cdots < m_n < \cdots$ such that $x_n^{m_n} \to_H \theta$.

9. Linear operations. Let X and Y be two sets of elements. If to each $x \in X$ there is made to correspond an element $y \in Y$ then an operation $y = U(x)$ is defined. X will be called the domain and Y the range of the operation.

DEFINITION 3. If $y = U(x)$ is an operation which transforms X into all or a part of Y, then if

(1) X and Y are H-normed spaces $y = U(x)$ is said to be HH-continuous at the point x_0 if $x_n \to_H x_0$ implies $U(x_n) \to_H U(x_0)$,
(2) X is an H-normed space and Y is a linear normed space $y = U(x)$ is said to be HB-continuous at the point x_0 if $x_n \to_H x_0$ implies $U(x_n) \to_B U(x_0)$,
(3) X is a linear normed space and Y is an H-normed space $y = U(x)$ is said to be BH-continuous at the point x_0 if $x_n \to_B x_0$ implies $U(x_n) \to_H U(x_0)$,
(4) X and Y are linear normed spaces $y = U(x)$ is said to be BB-continuous at the point x_0 if $x_n \to_B x_0$ implies $U(x_n) \to_B U(x_0)$.

If in any one of the above cases the specified continuity condition holds for every point of X then the operation is said to be continuous in that sense on X.

THEOREM 9.1. *If X is an H^*-normed space, Y is an H-normed space, and $y = U(x)$ is HH-continuous, then it is BH-continuous.*

PROOF. $x_n \to_B x$ implies $x_n \to_H x$ in an H^*-normed space, so by the HH-continuity of $y = U(x)$, $x_n \to_B x$ implies $U(x_n) \to_H U(x)$.

THEOREM 9.2. *If X is a linear normed space, Y is an H^*-normed space, and $y = U(x)$ is BB-continuous, then it is BH-continuous.*

THEOREM 9.3. *If X is an H-normed space, Y is an H^*-normed space, and $y = U(x)$ is HB-continuous then it is HH-continuous.*

THEOREM 9.4. *If the additive operation $y = U(x)$ transforms an H-normed space X into a space Y of the same type and is HH-continuous at a single point, then it is HH-continuous on X.*

THEOREM 9.5. *Statements similar to that of the previous theorem hold when $y = U(x)$ is HB-, BH-, or BB-continuous at a single point, X is H-normed, linear normed, or linear normed, respectively, and Y is linear normed, H-normed, or linear normed, respectively.*

THEOREM 9.6. *If $y = U(x)$ is an HH-continuous and additive operation from an $\bar{H}$-normed space X to an $\bar{H}$-normed space Y, then it is homogeneous.*

PROOF. The proof of this theorem depends on the uniqueness of the H-limit notion in $\bar{H}$-normed spaces.

THEOREM 9.7. *Statements similar to that of the previous theorem hold when $y = U(x)$ is additive and HB-, BH-, or BB-continuous, X is $\bar{H}$-normed, linear normed, or linear normed, respectively, and Y is linear normed, $\bar{H}$-normed, or linear normed, respectively.*

In case the operation is from a domain X to the set of all real numbers, which set is readily seen to be an $\bar{H}$-normed space when the norm of §1 is used, it will be called a functional and will be designated by $y = f(x)$. Since in the $\bar{H}$-normed space of real numbers (see §1) the H- and B-convergence notions are equivalent (see Theorem 1.1), the four types of operations of Definition 3 reduce to two in the case of functionals. These will be referred to as the H- and B-continuous functionals.

10. General forms of H-continuous and linear functionals. Remark 10.1. Let X be an H^*-normed space. By Theorem 9.1 every H-continuous functional

on X is B-continuous on X. Consequently, the general form of the H-continuous and linear functional on X must be of the same nature as but less general than the general form of the B-continuous and linear functional on X.

10.1. The space (cH). The elements of this space form a Banach space (c) when the norm of a finite complex operates as a norm for single elements. Then (cH) is an H^*-normed space. The general form of the B-continuous and linear functional for the space (c) is $f(x) = C \lim_{i=\infty} \xi_i + \sum_{i=1}^{\infty} C_i\xi_i$ (Banach, p. 66).

THEOREM 10.1. *The general form of the H-continuous and linear functional on (cH) is* $f(x) = C \lim_{i=\infty} \xi_i + \sum_{i=1}^{n} C_i\xi_i$.

PROOF. By Remark 10.1 every H-continuous and linear functional on (cH) is of the form of the B-continuous and linear functional on the same elements taken as a Banach space but with additional restrictions on the constant C_i's. Assume that $f(x)$ is H-continuous and linear and that there is an infinite number of non-zero constant C_i's involved in its representation. Consider the sequence $\{x_n\}$ of (cH) where $\xi_{ni} = 0$ if $n \neq i$ and $\xi_{nn} = \operatorname{sgn} 1/C_n$. This sequence is obviously such that $x_n \rightarrow_H \theta$. But for an infinite number of values of $n, f(x_n) = 1$, so $\overline{\lim_{n=\infty}} f(x_n) = 1$. Consequently, $x_n \rightarrow_H \theta$ does not imply $f(x_n) \rightarrow_B 0$. Therefore, the possibility of the general form of the H-continuous and linear functional on (cH) involving more than a finite number of non-zero constants has been eliminated. Consequently, we are confined to a consideration of the form $C \lim_{i=\infty} \xi_i + \sum_{i=1}^{n} C_i\xi_i$. This functional is readily seen to be H-continuous and linear.

10.2. The space (l^pH), $p \geqq 1$. Remarks similar to those preceding Theorem 10.1 apply here.

THEOREM 10.2. *The general form of the H-continuous and linear functional on (l^pH), $p \geqq 1$, is* $f(x) = \sum_{i=1}^{n} C_i\xi_i$.

PROOF. The proof of this theorem is similar to the proof of Theorem 10.1.

REMARK 10.2. If K is a finite complex then by $U(K)$ is meant the complex of the images of the elements of K by $y = U(x)$ (Vulich 1, p. 165). Vulich has proved the following theorem for operations defined on K-formed spaces: A necessary and sufficient condition that the additive operation $y = U(x)$ be KK-continuous is that there exist a constant $C > 0$ such that for every complex K of the domain $\| U(K) \| \leqq C \| K \|$. A similar statement cannot be made for HH-continuous and additive operations on H^*-normed spaces. For in the space (l^pH), $p \geqq 1$, let K be the complex (x_1, x_2) where $\xi_{pi} = 0$ if $p \neq i$ and $\xi_{pp} = 1$, and let $f(x) = \xi_1 + \xi_2$. Then $\| (x_1, x_2) \| = 0$ and $\| f(x_1), f(x_2) \| = 1$. Hence, there exists no constant $C > 0$ such that $\| f(K) \| \leqq C \| K \|$.

10.3. The space (CH). The elements of (CH) form a Banach space (C) when the norm of a finite complex operates on single elements only. Then (CH) is an H^*-normed space. The general form of the B-continuous and linear functional defined on the elements of (CH) when they are taken as the Banach space

(C) is $f(x) = \int_0^1 x(t)\,dg(t)$, where $g(t)$ is a function of bounded variation and $x(t)$ is of (C) (Banach, p. 61).

THEOREM 10.3. *The general form of the H-continuous and linear functional on (CH) is* $f(x) = \int_0^1 x(t)\,dg(t)$, *where* $g(t)$ *is a function of bounded variation which is constant except for a finite number of discontinuities. An alternate form is* $f(x) = \sum_{i=1}^{n} C_i x(t_i)$.

PROOF. It will be shown first that $g(t)$, which must be of bounded variation because of Remark 10.1, cannot be continuous and non-decreasing (or non-increasing) on (0, 1) with $g(t) > g(0)$ (or $g(t) < g(0)$) for $t > 0$. Suppose that $g(t)$ is of this form. Define a sequence $\{x_n\}$ of continuous functions as follows:

$$x_n(t) = x_n(0) = 0 \text{ for } 2/n \leqq t \leqq 1,$$

$$x_n(1/n) = I^{-1}, \text{ where } I = \int_0^{1/n} t\,dg(t),^{8}$$

and $x_n(t)$ linear elsewhere. Obviously, $x_n \rightarrow_H \theta$. It will be shown that $\lim_{n\to\infty} \int_0^1 x_n(t)\,dg > 0$. Write $\int_0^1 x_n(t)\,dg = \int_0^{1/n} x_n(t)\,dg + \int_{1/n}^{2/n} x_n(t)\,dg$. On the interval $(0, 1/n)$, $x_n(t) = ntI^{-1}$. Then $\int_0^{1/n} x_n(t)\,dg = \int_0^{1/n} nI^{-1}t\,dg = nI^{-1}I = n$. On the other hand, on the interval $(1/n, 2/n)$, $x_n(t) = -nI^{-1}(t - 2/n)$, whence $\int_{1/n}^{2/n} x_n(t)\,dg = -nI^{-1}\int_{1/n}^{2/n} (t - 2/n)\,dg = -nI^{-1}\int_{1/n}^{2/n} t\,dg + 2I^{-1}\int_{1/n}^{2/n} dg$. Since in $(1/n, 2/n)$, $t \leqq 2/n$, it follows that $\int_{1/n}^{2/n} t\,dg \leqq 2/n \int_{1/n}^{2/n} dg$, and so $\int_{1/n}^{2/n} x_n(t)\,dg \geqq -2I^{-1}\int_{1/n}^{2/n} dg + 2I^{-1}\int_{1/n}^{2/n} dg = 0$. Consequently, $\int_0^1 x_n(t)\,dg \geqq n$, and therefore, if $g(t)$ is of the form specified at the beginning of this proof, $x(t)$ is of (CH), and $f(x) = \int_0^1 x(t)\,dg(t)$, then there is a sequence $\{x_n\}$ such that $x_n \rightarrow_H \theta$ but for which $\{f(x_n)\}$ does not approach zero as $n \rightarrow \infty$. Then $f(x)$ is not H-continuous. So, in general if $f(x)$ is H-continuous and linear it may not be of the form $\int_0^1 x(t)\,dg(t)$, where $g(t)$ is continuous and such that for some interval (t_0, t'), $g(t_0) < g(t)$ (or $g(t_0) > g(t)$) for $t > t_0$. Since if this were the case the sequence $\{x_n\}$ used above could be constructed on the interval (t_0, t') and then a contradiction would follow.

Therefore, $g(t)$ must be constant except for at most a denumerable number of points of discontinuity of the first kind. Assume that $g(t)$ is of this nature

---

[8] By the integration by parts theorem for Stieltjes integrals it is readily shown that the denominator does not vanish.

and that $f(x)$ is an H-continuous and linear functional on (CH). Let $\{a_n\}$ be a sequence of points of $(0, 1)$ which approaches some a_0 from the right (or left) and which has the property that at each a_n the function $g(t)$ has a discontinuity of measure $b_n \neq 0$. Each a_n may itself be the limit point of points of discontinuity of $g(t)$. Let such points be designated by a_{np} and let the measure of the discontinuity of $g(t)$ at a_{np} be $b_{np} \neq 0$. About each a_n choose an interval I_n such that the left-hand end point of I_n is to the right of a_0, such that the length of $I_n \leqq \min(\text{length } I_{n-1}, 1/n)$, and such that if $a_{np} \in I_n$ then $\sum_{p=1}^{\infty} | b_{np} | \leqq | b_n |/2^n$. Define a sequence of continuous functions as follows: let $x_n(a_n) = \operatorname{sgn} 1/b_n$; on the complement of I_n let $x_n(t) = 0$, and finally let $x_n(t)$ be linear elsewhere. Then $x_n \rightarrow_H \theta$. $\int_0^1 x_n(t)\, dg = \sum_{p=1}^{\infty} [g(a_{np} + 0) - g(a_{np} - 0)]x_n(a_{np}) + b_n x_n(a_n) = \sum_{p=1}^{\infty} b_{np} x_n(a_{np}) + 1$. If each $b_{np} x_n(a_{np}) < 0$, then $\int_0^1 x_n(t)\, dg = 1 - \sum_{p=1}^{\infty} | b_{np} || x_n(a_{np}) | \geqq 1 - \sum_{p=1}^{\infty} | b_{np} |/2^n | b_n | > 1/2$. If each $b_{np} x_n(a_{np})$ is not less than zero then $\int_0^1 x_n(t)\, dg$ is obviously greater than $1/2$. Therefore, although $x_n \rightarrow_H \theta$ it does not follow that $f(x_n) \rightarrow_B 0$. Hence, if $f(x)$ is to be H-continuous and linear it is impossible that $g(t)$ be other than constant except for at most a finite number of points of discontinuity of the first kind. In this case $f(x) = \int_0^1 x(t)\, dg = \sum_{i=1}^{n} x(t_i)[g(t_i + 0) - g(t_i - 0)] = \sum_{i=1}^{n} C_i x(t_i)$. Since only a finite number of terms are involved in this form $f(x)$ is obviously H-continuous and linear.

REMARK 10.3. It is evident from this theorem that there are B-continuous and linear functionals which are not H-continuous. Therefore, the converse of Theorem 9.1 is not true.

10.4. The space (L^pH), $p \geqq 1$. THEOREM 10.4. *If $y = U(x)$ is a continuous transformation from (L^pH), $p \geqq 1$, to a space Y in which the limit notion is unique, then $y = U(x)$ is a constant.*[9]

PROOF. Let $x \in (L^pH)$, $p \geqq 1$, and let $y_n(t) = x(t) \cdot 1/2(1 + r(t))$, where $r_n(t)$ is the function defined in Remark 3.1. As in Remark 3.1 it is seen that $y_n \rightarrow_H \theta$ and $y_n \rightarrow_H x$. Since $y = U(x)$ is continuous it follows that $U(y_n) \rightarrow U(\theta)$ and $U(y_n) \rightarrow U(x)$. Because of the uniqueness of the limit notion in Y, $U(x) = U(\theta)$.

COROLLARY 10.41. *If $y = f(x)$ is an H-continuous and linear functional from (L^pH), $p \geqq 1$, then $f(x) \equiv 0$.*

11. General forms of different types of linear operations. If for the set of all continuous functions defined on $(0, 1)$ the norm for single elements is taken to be $\max_{0 \leqq t \leqq 1} | x(t) |$, then the set is a Banach space and it will be denoted by (C). With respect to this norm the space (CH) is an H^*-normed space. It has been

[9] This theorem and its proof were supplied by the referee. It is a generalization of a theorem originally given by the author.

shown (Fichtenholz, p. 32) that the general form of the BB-continuous and linear operation from (C) to (C) is given by

$$(1) \qquad y(s) = U(x) = \int_0^1 x(t)\, d_t K(s, t),$$

where $K(s, t)$ is a function of two variables defined on $0 \leqq s, t \leqq 1$ and is such that

(1^0) $K(s, t)$ is of bounded variation as a function of t for each s and $\operatorname{var}_{0 \leqq t \leqq 1} K(s, t) \leqq C$, where C is a positive constant independent of s,

(2^0) $K(s, 0) = 0$; $K(s, 1)$ is continuous in s,

(3^0) $K(s, t)$ is continuous in measure[10] with respect to s in the interval $(0, 1)$ for each $s = s_0$.

If $y(s) = U(x)$ is BH-continuous and linear from (CH) to (CH), then $y(s_0) = U(x)$ is a B-continuous and linear functional defined on (CH) and it may be represented by $y(s_0) = U(x) = \int_0^1 x(t)\, d_t K(s_0, t)$, and so $y(s) = U(x)$ is of form (1) with restrictions on $K(s, t)$ which will now be examined.

THEOREM 11.1. *The class of all BH-continuous and linear operations from (CH) to (CH) is equivalent to the class of all BB-continuous and linear operations from (CH) to (CH).*

PROOF. Because of Theorem 9.2 it is evident that the class of all BB-continuous and linear operations from (CH) to (CH) is a subclass of the class of all BH-continuous and linear operations between the same spaces. Therefore, since each BH-continuous and linear operation is of the form (1), in order that an operation be BH-continuous and linear without being BB-continuous and linear it is necessary that some one of the conditions 1^0, 2^0, 3^0 on $K(s, t)$ be weakened. But these conditions are necessary and sufficient not only for the BB-continuity and linearity of $y = U(x)$ but also for the continuity of $y(s) = U(x)$. Hence, if one of these conditions is weakened $y(s)$ will no longer be continuous. Therefore, every BH-continuous and linear operation must satisfy conditions 1^0, 2^0, and 3^0 and our theorem is proved.

THEOREM 11.2. *The general form of the HH-continuous and linear operation from (CH) to (CH) is given by* (1) *where $K(s, t)$ satisfies conditions 1^0, 2^0, 3^0 and also*

(4^0) *for s constant $K(s, t)$ is constant in t except for at most a finite number of points of discontinuity of the first kind.*

PROOF. If $y(s) = U(x)$ is of form (1) and if $K(s, t)$ satisfies conditions 1^0-4^0, then $y(s)$ is continuous because of conditions 1^0-3^0. Moreover, if $x_n \to_H x$, then $y_n(s) \to y(s)$ for each $s \in (0, 1)$ and so $y_n \to_H y$. For consider any $s = s_0$, then $| y_n(s_0) - y(s_0) | \leqq \left| \int_0^1 [x_n(t) - x(t)]\, d_t K(s_0, t) \right|$. The last expression is the

[10] $K(s, t)$ is said to be convergent in measure with respect to s for $s = s_0$ on $0 \leqq t \leqq 1$ if $K(s, t)$ as a function of t approaches $K(s_0, t)$ in measure when $s \to s_0$.

absolute value of an H-continuous and linear functional (see Theorem 10.4 and condition 4^0), and therefore, $x_n \rightarrow_H x$ implies $y_n(s_0) \rightarrow y(s_0)$. This proves our contention.

If $y(s) = U(x)$ is an HH-continuous and linear operation from (CH) to (CH), then it is of form (1) and $K(s, t)$ satisfies 1^0–4^0. For consider any $s = s_0$, then $y(s_0) = U(x)$ is merely an H-continuous and linear functional on (CH). Hence, $y(s_0) = \int_0^1 x(t)\, d_tK(s_0, t)$ and $K(s_0, t)$ must satisfy 4^0 because of the restriction of Theorem 10.4. Then $y(s) = \int_0^1 x(t)\, d_tK(s, t)$. Since $y(s) = U(x)$ is HH-continuous and linear and (CH) is an H^*-normed space, it follows from Theorem 9.1 that $y(s)$ is BH-continuous and linear. By Theorem 11.1, $y(s) = U(x)$ is also BB-continuous and linear and so $K(s, t)$ must satisfy conditions 1^0–3^0.

THEOREM 11.3. *The general form of the HB-continuous and linear operation from* (CH) *to* (CH) *is given by* (1) *where* $K(s, t)$ *satisfies conditions* 1^0–3^0 *and also*

(5^0) *there exists at most a finite number of points* $t_1, t_2, \cdots, t_q$ *such that for any value* $s = s_0$ *the points of discontinuity of* $K(s_0, t)$ *are to be found among the points* $t_i (i = 1, 2, \cdots, q)$.

PROOF. As in the proof of Theorem 11.2 it is readily seen that any HB-continuous and linear operation is of the form (1). If $y = U(x)$ is HB-continuous and linear, then by Theorem 9.3 it is HH-continuous and linear, and so by Theorem 11.2 $K(s, t)$ satisfies conditions 1^0–3^0. Suppose that 5^0 is not satisfied by $K(s, t)$. Then there exists a sequence $\{s_n\}$ of points of the interval $(0, 1)$ such that the points of discontinuity of the first kind which are contributed by $\{K(s_n, t)\}$ are infinite in number, that is to say, there is an infinite sequence $\{t_\nu\}$ of points such that each is a point of discontinuity of some $K(s_n, t)$. Choose a subsequence $\{t_{\nu_p}\}$ of $\{t_\nu\}$ such that $\{t_{\nu_p}\}$ approaches some t_0 from the right (or left). About each t_{ν_p} as center and associated with each $K(s_n, t)$ there is an interval $I(t_{\nu_p}, s_n)$ such that within it there is no point of discontinuity of $K(s_n, t)$ except perhaps t_{ν_p} itself. The existence of this interval is assured by the HH-continuity of each HB-continuous operation. Let I_{ν_p} be an interval $I(t_{\nu_p}, s_n)$ which does not contain t_0 and whose right-hand end point is to the left of $t_{\nu_p} + (t_{\nu_p} - t_0)/2$ and which is such that the measure $a^n_{\nu_p}$ of the discontinuity of $K(s_n, t)$ at t_{ν_p} is not equal to zero. Define a sequence $\{x_n\}$ of continuous functions as follows:

$$x_n(t_{\nu_p}) = \operatorname{sgn} 1/a^n_{\nu_p},$$

$$x_n(t) = 0 \text{ for } t \text{ not of } I_{\nu_p},$$

and $x_n(t)$ linear elsewhere. Then $x_n \rightarrow_H \theta$. Moreover, $y_n(s) = \int_0^1 x_n(t)\, d_tK(s, t)$ is such that $| y_n(s_n) | = | x_n(t_{\nu_p}) | \cdot | K(s_n, t_{\nu_p} + 0) - K(s_n, t_{\nu_p} - 0) | = | \operatorname{sgn} 1/a^n_{\nu_p} | \cdot | a^n_{\nu_p} | = 1$. Therefore, $\| y_n \| = \max_{0 \leq s \leq 1} | y_n(s) | = 1$, and so $x_n \rightarrow_H \theta$ does not imply $y_n \rightarrow_B \theta$. This is a contradiction and so 5^0 must hold.

Let $y(s) = U(x)$ be of the form (1) and be defined on (CH), and let $K(s, t)$ satisfy conditions 1^0, 2^0, 3^0, and 5^0. Because of conditions 1^0–3^0, $y(s)$ is continuous. Moreover, $y(s) = \int_0^1 x(t)\, d_t K(s, t) = \sum_{i=1}^q x(t_i)[K(s, t_i + 0) - K(s, t_i - 0)] = \sum_{i=1}^q x(t_i)C_i(s)$ by condition 5^0. From 1^0 we know that there exists a constant $C > 0$ such that $| C_i(s) | \leqq C$ for each i and for $s \in (0, 1)$. Then $| y_n(s) - y(s) | \leqq C \sum_{i=1}^q | x_n(t_i) - x(t_i) |$. If $x_n \to_H x$, then since only a finite number of points are involved in the last summation, $|| y_n - y || \to 0$.

PHILADELPHIA COLLEGE OF PHARMACY

BIBLIOGRAPHY

S. Banach, *Théorie des Opérations Linéaires* (1932).

C. Caratheodory, *Vorlesungen über Reelle Funktionen* (1918).

G. Fichenholz, *Sur les opérations linéaires dans l'espace des fonctions continues*, Bull. Acad. Roy. Belg. (1936), v. 22, pp. 26-33.

B. Vulich 1, *On a generalized notion of convergence in a Banach space*, Annals of Math., v. 38, pp. 156-174.

B. Vulich 2, *Some remarks to the theory of K-normed spaces*, Comptes Rendus de l'Académie des Sciences de l'U.R.S.S. (1936), v. 2, #2, pp. 59-62.

B. Vulich 3, Correction to the note "*Some remarks to the theory of K-normed spaces,*" Comptes Rendus de l'Académie des Sciences de l'U.R.S.S. (1936), v. 3, #3, pp. 109-110.

Annals of Mathematics
Vol. 42, No. 1, January, 1941

CHARACTERIZATION OF COMPLEX COUPLE SPACES

By Aristotle D. Michal and Max Wyman

(Received February 13, 1940)

Introduction. In considering generalisations of classical complex analysis for complex Banach spaces it seems desirable to consider complex couple spaces. For such spaces many of the classical theorems can be shown to hold.[1] The following paper attempts to characterize such spaces. This is done by means of a conjugate operation.

1. Characterization of Hermitian couple spaces.

Definition 1.1. Two complex Banach spaces B_1, B_2 shall be called equivalent[2] if there exists a single valued function $f(z)$ on B_1 to B_2 with the properties

(a) $f(z)$ is linear[3] in z, and is homogeneous of degree one,

(b) $f(z)$ ranges over the whole of B_2, and is solvable in z,

(c) $\|f(z)\| = \|z\|$.

If conditions (a), (b) are satisfied we say B_1, B_2 are linearly homeomorphic.

Let E be a real Banach space with a real inner product $[x, y]$[4] From E we can construct a complex Banach space $E(c)$ as follows. Let $E(c)$ be the set of all couples $\{x, y\}$ where x, y are elements of E. We define

(1) $\{x_1, y_1\} = \{x_2, y_2\}$ if and only if $x_1 = x_2$, $y_1 = y_2$,

(2) $\{x_1, y_1\} + \{x_2, y_2\} = \{x_1 + x_2, y_1 + y_2\}$,

(3) $(a + ib)\{x, y\} = \{ax - by, bx + ay\}$ where a, b are real numbers,

(4) $\|\{x, y\}\| = (\|x\|^2 + \|y\|^2)^{\frac{1}{2}}$

Definition 1.2. A complex Banach space B is called a Hermitian couple space, if it is equivalent to a space of type $E(c)$.

Definition 1.3. Let B be a complex Banach space. A bilinear function $[Z, U]$ on B^2 to the complex numbers is called a Hermitian inner product[5] if

(i) $\overline{[Z, U]} = [U, Z]$, $[(a + ib)Z, U] = (a + ib)[Z, U]$,

[1] For work in general analysis see Taylor (3), Michal and Martin (4), and Michal, Davis and Wyman (6).

[2] For the notion of equivalence of real Banach spaces see Banach (1), p. 180.

[3] We use the term linear to mean additive and continuous. In real Banach spaces this is enough to ensure homogeneity of degree one. This is no longer so in complex Banach spaces.

[4] For postulates of a real inner product see Michal (5). In addition we make the restriction that $\|x\| = [x, x]^{\frac{1}{2}}$.

[5] These are of course the same postulates given for an inner product in Hilbert space. See Stone (2) and the references to Von Neumann's papers given there.

(ii) $[Z, Z] \geqq 0$ and $[Z, Z] = 0$ if and only if $Z = 0$,
(iii) $\| Z \| = [Z, Z]^{\frac{1}{2}}$.

THEOREM 1. *A necessary and sufficient condition that an arbitrary complex Banach space B be a Hermitian couple space is that*
(α) *B must possess a Hermitian inner product* $[Z, U]$
(β) *there exist a function* $\bar{Z}$[6] *on B to B such that*

$$(\beta_1) \quad \overline{[Z_1, Z_2]} = [\bar{Z}_1, \bar{Z}_2],$$

$$(\beta_2) \quad \bar{\bar{Z}} = Z.$$

Proof of necessity. For a space $E(c)$ of couples $z = \{x, y\}$ the existence of a Hermitian inner product[7] $[z, u]$ was shown in Michal, Davis and Wyman (6). The function $\bar{z}$ for such spaces is the "ordinary complex" conjugate $\bar{z} = \{x, -y\}$. Let $Z = f(z)$ be the mapping function which makes B equivalent to $E(c)$. We define
(a) $[Z, U] = [f^{-1}(Z), f^{-1}(U)]$, where the inner product on the right is that of $E(c)$,
(b) $\bar{Z} = f(\bar{z})$.
By straightforward calculation $[Z, U]$ is a Hermitian inner product for B, and condition (β) is satisfied. To prove the sufficiency we make use of the following lemmas.

LEMMA 1. *Let B be a complex Banach space satisfying* (α), (β) *of theorem* 1. *Then* $\bar{Z}$ *is linear in* Z, *and*
(1.1) $\overline{(a + ib)Z} = (a - ib)\bar{Z}$, *for a, b real.*
For all Z_1, Z_2, Z_3 we have

$$(1.2) \quad \overline{[\overline{Z_1 + Z_2}, Z_3]} = \overline{[Z_1 + Z_2, \bar{Z}_3]} = [\bar{Z}_1, Z_3] + [\bar{Z}_2, Z_3] = [\bar{Z}_1 + \bar{Z}_2, \bar{Z}_3].$$

Thus $\overline{Z_1 + Z_2} = \bar{Z}_1 + \bar{Z}_2$, and $\bar{Z}$ is additive in Z. From $\| \bar{Z} \| = \| Z \|$ we can conclude that $\bar{Z}$ is linear in Z. Similarly property (1.1) is shown.

LEMMA 2. *The totality E of all elements W for which* $\bar{W} = W$ *is a real sub-Banach space of B and is such that* $\| W \|$ *is generated by a real symmetric inner product.*

Clearly under the operations of B, E is a real normed linear space. If $\{W_i\}$ is a Cauchy convergent sequence of E there is an element W of B such that $\| W - W_n \| < \frac{\epsilon}{2}$. From $\| \bar{Z} \| = \| Z \|$ we obtain $\| \bar{W} - W_n \| < \frac{\epsilon}{2}$, and hence $\| W - \bar{W} \| < \epsilon$. Thus $W = \bar{W}$ and W is an element of E. Thus E is a complete space, and hence is a real Banach space.

---

[6] We have used the notation $\bar{Z}$ as a function on B to B. This should not be confused with the complex conjugate notation of ordinary complex numbers. See (β_1) where both occur.

[7] If x, y, ζ, η are elements of E, and $z = \{x, y\}$, $u = \{\zeta, \eta\}$ are in $E(c)$, then the Hermitian inner product $[z, u]$ of $E(c)$ is given by $[z, u] = [x, \zeta] + [y, \eta] + i\{[y, \zeta] - [x, \eta]\}$, where $[x, y]$ is the inner product of E.

For any W_1, W_2 of E we have

$$(1.3) \qquad \overline{[W_1, W_2]} = [\overline{W}_1, \overline{W}_2] = [W_1, W_2] = [W_2, W_1].$$

This completes the proof of lemma 2.

From the E of lemma 2, let us construct the Hermitian couple space $E(c)$. For any element Z of B, $\left\{\frac{Z + \overline{Z}}{2}, \frac{i\overline{Z} - iZ}{2}\right\}$ is an element of $E(c)$. Then B is equivalent to $E(c)$ and the required mapping function is

$$(1.4) \qquad f(Z) = \left\{\frac{Z + \overline{Z}}{2}, \frac{i\overline{Z} - iZ}{2}\right\}$$

The additivity and homogeneity of degree one of $f(Z)$ are easy to verify. From $\| Z \|^2 = [Z, Z]$ we readily obtain that

$$\| f(Z) \| = \left(\frac{\| Z + \overline{Z} \|^2}{4} + \frac{\| i\overline{Z} - iZ \|^2}{4}\right)^{\frac{1}{2}} = \| Z \|.$$

Finally the equation $f(Z) = \{W_1, W_2\}$ has the unique solution $Z = W_1 + iW_2$ for any W_1, W_2 of E. Thus B is a Hermitian couple space.

2. Characterization of complex couple spaces. If E is an arbitrary Banach space, the couple space $E(c)$ is a complex Banach space under definitions (1)–(4) if and only if the norm of E is generated by a real inner product. In a previous paper (Michal, Davis and Wyman (6)) the authors raised the question of giving a definition to replace (4) which would eliminate the restriction on the norm of E. A definition was communicated to us by E. W. Paxson which could be used if E were separable. Later on A. E. Taylor[8] gave us verbally a definition which places no restriction on E. This result is stated in the following theorem.

THEOREM 2. *Let E be an arbitrary real Banach space, and let x, $y \in E$. The totality $E'(c)$ of couples $\{x, y\}$ form a complex Banach space under definitions* (1), (2), (3) *and*

$$(4') \qquad \| \{x, y\} \| = \sup_{\text{mod } l=1} (l^2(x) + l^2(y))^{\frac{1}{2}}.\text{[9]}$$

The truth of this theorem is easily verified. Complex spaces which are linearly homeomorphic to spaces of the type $E'(c)$ are called complex couple spaces.

THEOREM 3. *A necessary and sufficient condition that an arbitrary complex Banach space B be a complex couple space is that there exist a function $\overline{Z}$ on B to B with the properties*:

$$\text{I} \qquad \overline{Z_1 + Z_2} = \overline{Z}_1 + \overline{Z}_2,$$

[8] Both Taylor and Paxson have seen the manuscript of Michal, Davis and Wyman (6). Taylor had also seen Paxson's result.

[9] $l(x)$ means a linear functional on E to the real numbers. The superior is taken over all linear functionals whose modulus is one.

II $$\overline{iZ} = -i\bar{Z},$$

III $$\bar{\bar{Z}} = Z,$$

IV $$\| \bar{Z} \| = \| Z \|.$$

The proof follows essentially the same lines as the proof of theorem 1. $\bar{Z}$ will now trivially satisfy lemma 1, and the first part of lemma 2 will be satisfied. From E we construct $E'(c)$ and the latter will be linearly homeomorphic to B. The continuity of the mapping function $f(Z) = \left\{\frac{Z + \bar{Z}}{2}, \frac{i\bar{Z} - iZ}{2}\right\}$ follows from $\| f(Z) \| \leqq \sqrt{2} \, \| Z \|$.

CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA.

BIBLIOGRAPHY

(1) Banach—*Théorie des Opérations Linéaires.*
(2) Stone—*Linear Transformations in Hilbert Space.*
(3) A. E. Taylor—*Analytic functions in general analysis*—Annali Della R. Scuola Normali Di Pisa, Ser. 2, vol. 6 (1937), pp. 277–291.
(4) A. D. Michal and R. S. Martin—*Some expansions in vector space*—Journal de Mathematique, Serie 9, vol. 13–14 (1934), pp. 69-91.
(5) A. D. Michal—*Abstract covariant vector fields in a general absolute calculus*—American Journal of Mathematics, vol. 59 (1937), pp. 306-314.
(6) A. D. Michal, R. Davis and M. Wyman—*Polygenic functions in general analysis*—Annali Della R. Scuola Normali Di Pisa—In press.

Annals of Mathematics
Vol. 42, No. 1, January, 1941

SEQUENCES DEFINED BY NON-LINEAR ALGEBRAIC DIFFERENCE EQUATIONS[1]

By Otis E. Lancaster

(Received March 7, 1938)

An algebraic difference equation

$$\Phi(y(x + m), y(x + m - 1), \cdots, y(x), x) = 0, \tag{1}$$

Φ a polynomial with rational coefficients in its arguments x, $y(x)$, $y(x + 1)$, $\cdots$, $y(x + m)$, defines a sequence[2]

$$y(a), y(a + 1), \cdots, y(a + m - 1), y(a + m), \cdots$$

when the values of $y(x)$ are assigned at the m points $a, a + 1, \cdots, a + m - 1$. Obviously every equation (1) defines an infinite number of sequences: one for each set of values $y(a)$, $y(a + 1)$, $\cdots$, $y(a + m - 1)$.

We propose to obtain some of the properties of such sequences by studying the difference equations which define them. In the study of infinite sequences one is mainly interested in their ultimate behavior. Hence we are here interested in the behavior of the solutions of the difference equations, for integral values of x, in the neighborhood of infinity. Although, at present, it is not possible to write out explicitly the solutions of most non-linear difference equations, it seems that it might be feasible to determine whether a given equation defines sequences that approach zero as x becomes infinite by considering the solutions of the linear difference equation formed by omitting the non-linear terms. (We assume that at least one linear term appears.) For when $y(x)$ approaches zero the linear terms are infinitesimals of the first order, while the non-linear terms are infinitesimals of higher order. Therefore, it would be expected that the behavior of such sequences is largely determined by the linear terms of the difference equation. With this idea in mind, we attempt to gain information about the sequences defined by a non-linear difference equation by considering the solution of the difference equation formed from its linear terms.[3]

It is clear that once we have criteria for determining when a difference equation defines sequences which approach zero we can easily determine whether one defines sequences that approach a constant limit α. For, if after the application of the transformation $y(x) = \bar{y}(x) + \alpha$, the transformed difference equa-

---

[1] The author's thanks are due to Professor G. D. Birkhoff for his counsel during the preparation of this paper.

[2] When the equation (1) is non-linear in the variable $y(x + m)$ some convention must be made as to the one of the values of $y(x + m)$ that is to be taken at each step.

[3] The investigations of Lattès (1), Horn (2) and Perron (3) have also had as their basis the same idea.

tion defines sequences that approach zero, then the original equation defines sequences that approach α.

Besides determining conditions sufficient to insure that the sequences of a difference equation converge to a definite limit, we study the effect of the degree of the equation, of the order of the equation, and of variable coefficients upon the rate of convergence of the sequences. Lastly we determine difference equations whose rational sequences converge most rapidly to irrational numbers of the form $\beta^{1/p}$, where p is an integer and β is a rational number.

For the discussion we write (1) in the form

$$\Phi(S_{n+m}, S_{n+m-1}, \cdots, S_n, n) = 0, \tag{2}$$

where S_{n+i} denotes the value of $y(n+i)$ when n is an integral value of x. The sequences are then denoted by $\{S_n\}$. Of course, when we speak of the solutions of the equation (2) it is understood that n is a continuous real variable, and $y(n) = S_n$.

1. Sequences that approach zero.

DEFINITION. The phrase "*the difference equation of the linear terms*" shall denote the difference equation formed from the given equation by omitting all terms that are of higher than the first degree in the arguments S_n, S_{n+1}, $\cdots$, S_{n+m}, where it is understood that the linear equation is to be of the *same order* as that of the non-linear equation.

For example, the difference equation of the linear terms of the non-linear equation,

$$S_{n+2}S_nS_{n+1} + nS_nS_{n+2} + S_n + nS_{n+1} + n^2 = 0$$

is

$$0\cdot S_{n+2} + nS_{n+1} + S_n = -n^2.$$

The characteristic equation of this linear equation is

$$0\cdot\rho^2 + 1\rho + 0 = 0.$$

It follows from the above definition that besides the usual possibilities for infinite and zero roots of the characteristic equation, the characteristic equation of the equation of the linear terms always has an infinite root if S_{n+m} does not appear linearly in the non-linear equation and a zero root if S_n does not appear linearly.

THEOREM 1. *If in a non-linear algebraic difference equation a coefficient of a linear term is of as high degree in n as the coefficients of the non-linear terms, and if there are solutions of the difference equation of linear terms which approach zero as n becomes infinite, and if the characteristic equation of the difference equation of linear terms has finite, non-zero roots not equal to one in absolute value, then there are sequences defined by the given equation that approach zero as n becomes positively infinite.*

The truth of this theorem follows immediately from the following theorem of Horn[2]. "In the system of non-linear difference equations

$$(3) \qquad y_i(x+1) = G_i(x, y_1(x), \cdots, y_m(x)), \qquad (i = 1, 2, \cdots, m)$$

let G_i be a function of $x, y_1(x), \cdots, y_m(x)$, which vanishes at the point $x = \infty$, $y_1 = 0, \cdots, y_m = 0$ and is analytic in a neighborhood of it:

$$G_i(x, y_1(x), \cdots, y_m(x)) = \frac{a_{i0}}{x} + a_{i1}y_1 + \cdots + a_{im}y_m + \cdots \qquad (i = 1, \cdots, m).$$

If $s - a_1, \cdots, s - a_m$ are the elementary divisors of the determinant

$$| a_{ij} - s\delta_{ij} |,$$

let the system be transformed to the system

$$(4) \qquad y_i(x+1) = a_i y_i(x) + f_i(1/x, y_1(x), \cdots, y_m(x)) \qquad (i = 1, \cdots, m)$$

where

$$f_i(1/x, y_1, \cdots, y_m) = \sum A^{(i)}_{\lambda,\lambda_1,\lambda_2,\cdots,\lambda_m}(1/x)^{\lambda} y_1^{\lambda_1} \cdots y_m^{\lambda_m}$$
$$(\lambda = 1, \lambda_1 = \lambda_2 = \cdots = \lambda_m = 0: \lambda + \lambda_1 + \cdots + \lambda_m \geqq 2)$$

and

$$0 < | a_1 | \leqq | a_2 | \leqq \cdots \leqq | a_\mu | < 1 < | a_{\mu+1} | \leqq \cdots \leqq | a_m |.$$

If it is possible to choose h so small and x_0 so large that

$$| f_i(1/x, y_1, \cdots, y_m) | < hq$$

and

$$| f_i(1/x, y_1', \cdots, y_m') - f_i(1/x, y_1, \cdots, y_m) |$$
$$(5) \qquad \leqq g(| y_1' - y_1 | + \cdots + | y_m' - y_m |) \text{ for } x > x_0, | y_i | \leqq h, | y_i' | \leqq h$$
$$(i = 1, 2, \cdots, m),$$

where $q \leqq \frac{1}{2} | a_i |^{\frac{1}{2}}(1 - | a_i |^{\frac{1}{2}})$, $(i = 1, 2, \cdots, \mu)$; $q \leqq | a_i |^{\frac{1}{2}}(| a_i |^{\frac{1}{2}} - 1)$ $(i = \mu + 1, \cdots, m)$, $g < q/m$ and if $\eta_i \leqq h/2$, *then* there exists a unique solution[4] $y_i(x)$, $(i = 1, \cdots, m)$ of the system for which $| y_i | \leqq h$ for $x > x_0$, $y_i(x_0) = \eta_i$ $(i = 1, 2, \cdots, \mu)$ and $\lim_{x\to\infty} y_i(x) = 0$."[5]

---

[4] In order to obtain a unique solution Horn admitted only constants for the arbitrary periodic functions of period one. The theorem holds when any periodic functions are admitted if the η_i are defined over the interval $(x_0, x_0 + 1)$.

[5] Although the theorem above is "translated" directly from Horn's paper (2), we see that if two or more of the a's are equal then the system (3) cannot in general be brought to the form (4). Since, however, an obvious extension of his work shows that the theorem holds for the case of equal a's, we assume the theorem for the general case. If some of the a's are equal the equations (4) are of the form

$$y_i(x+1) = a_i y_i + k y_{i-1}(x) + f(1/x, y_1, \cdots, y_m)$$

when $a_i = a_{i-1}$.

We prove Theorem 1 by showing that the conditions of Horn's theorem are all fulfilled. Since, by hypothesis, a solution of the equation of the linear terms approaches zero as $n \to \infty$, $\Phi(S_{n+m}, \cdots, S_n, n)$ vanishes at the point $n = \infty$, $S_n = 0, \cdots, S_{n+m} = 0$. Hence (2) may be written in the form

$$\begin{aligned}(6)\qquad & a_m S_{n+m} + a_{m-1} S_{n+m-1} + \cdots + a_0 S_n \\ & \qquad + \Phi_1(S_{n+m}, S_{n+m-1}, \cdots, S_n, n) + a_{-1} n^{-1} = 0,\end{aligned}$$

where Φ_1 is a polynomial whose terms are of higher than first degree in the arguments $1/n$, $S_n, \cdots, S_{n+m}$. It follows from the assumption that the roots of the characteristic equation of the equation of the linear terms are *finite*, that $a_m \neq 0$. And when $a_m \neq 0$, (2) treated as an algebraic equation in S_{n+m} has a solution

$$(7)\qquad S_{n+m} = f(S_{n+m-1}, \cdots, S_n, n)$$

which vanishes at the point $S_{n+m-1} = 0, \cdots, S_n = 0$, $n = \infty$. That is, the expansion of f in terms of $1/n$, $S_n, \cdots, S_{n+m-1}$ does not contain a constant term.

We show that equation (7) defines sequences that approach zero, from which it follows that equation (2) defines sequences which approach zero. Replacing S_{n+m} in Φ_1 by f, we have (7) in the form

$$\begin{aligned}(8)\qquad S_{n+m} = & -\frac{1}{a_m}[a_0 S_n + a_1 S_{n+1} + \cdots + a_{m-1} S_{n+m-1} \\ & + \Phi_1(f, S_{n+m-1}, \cdots, S_n, n) + a_{-1} n^{-1}];\end{aligned}$$

where the terms in the expansion of Φ_1 are of higher than first degree in $1/n$, $S_n, \cdots, S_{n+m-1}$. Upon setting $u_1(n) = S_n$, $u_2(n) = S_{n+1}, \cdots, u_m(n) = S_{n+m-1}$, we obtain from the equation (8) the system of difference equations

$$\begin{aligned} & u_i(n+1) = u_{i+1}(n) \qquad (i = 1, 2, \cdots, m-1) \\ (9)\qquad & u_m(n+1) = -\frac{1}{a_m}[a_0 u_1(n) + \cdots + a_{m-1} u_m(n) \\ & \qquad + \Phi_1(f, u_m(n), \cdots, u_1(n), n) + a_{-1} n^{-1}].\end{aligned}$$

The auxiliary equation of this system,

$$(10a)\qquad \begin{vmatrix} -\rho & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & -\rho & 1 \\ \frac{a_0}{a_m} & \frac{a_1}{a_m} & \frac{a_2}{a_m} & \frac{a_3}{a_m} & \cdots & \frac{a_{m-2}}{a_m} & \frac{a_{m-1}}{a_m} + \rho \end{vmatrix} = 0,$$

is equivalent to the characteristic equation of the equation of the linear terms, viz.

$$a_m\rho^m + a_{m-1}\rho^{m-1} + \cdots + a_0 = 0. \tag{10}$$

When the roots of (10) are distinct, we transform the system (9) into the canonical form

$$z_i(n+1) = \rho_i z_i(n) + \Psi_i(1/n, z_1(n), \cdots, z_m(n)) \qquad (i = 1, 2, \cdots, m) \tag{11}$$

and when (10) has equal roots, into one of the same form, except that when $a_j = a_{j-1}$ the j^{th} equation is

$$z_j(n+1) = \rho_j z_j(n) + z_{j-1}(n) + \Psi_j(1/n, z_1(n), \cdots, z_m(n)).$$

The Ψ_i $(i = 1, \cdots, m)$ satisfy the condition (5), since the terms in their expansions, except the one a_{-1}/n, are of higher than first degree in the variables $1/n$, $z_1(n), \cdots, z_m(n)$. Therefore, there are sequences $z_i(n)$ which approach zero as $n \to \infty$, and consequently, sequences $\{S_n\}$ defined by (7), that approach zero. Q.E.D.

It follows from the proof of Horn's theorem that an algebraic difference equation (linear or non-linear) cannot have more than one solution that approaches zero as $n \to \infty$ if the roots of the characteristic equation of the difference equation of the linear terms are greater than one in absolute value. Hence if there is only one solution of the difference equation of the linear terms that approaches zero as n becomes infinite there is at most one solution of the non-linear difference equation that approaches zero. However, if there are an infinite number of solutions of the difference equation of the linear terms which approach zero and if the roots of the characteristic equation are not equal to one in absolute value, then there are an infinite number of sequences of (7) that approach zero. Of course, it is clear that if a sequence is to approach zero the initial conditions S_{n_0+i} $(i = 0, \cdots, m-1)$, which determine it, must be chosen sufficiently small. Moreover, *if only one of the roots of the characteristic equation* (10) *is less than one in absolute value, then only one of these initial conditions is arbitrary.* If two roots are less than one in absolute value then two conditions are arbitrary, etc. If all the roots are less than one in absolute value, then all sequences defined by (7) converge to zero, for an arbitrary choice of initial conditions in the region $|S_{n_0+i}| \leqq h/2$, $(i = 1, \cdots, m-1)$. We call such a region a region of strong convergence.

When there is only one sequence defined by a non-linear difference equation that has a zero limit, it may be the trivial one $S_n = 0$ for all n. In fact, it is always such a sequence unless the equation has a term of zero$^{\text{th}}$ degree in the arguments S_n, $S_{n+1}, \cdots, S_{n+m}$. *In the future we exclude such sequences.*

The hypotheses of Theorem 1 exclude all the difference equations which do not contain both the linear terms a_0S_n and a_mS_{n+m}, $(a_0 \neq 0,\ a_m \neq 0)$. We have the following result concerning those in which $a_0 = 0$.

Theorem 2. *If in a non-linear algebraic difference equation a coefficient of a linear term is of higher degree in n than the coefficients of the non-linear terms, and all solutions of the difference equation of the linear terms approach zero as n becomes infinite, then there are sequences defined by the given equation that approach zero as* $n \to \infty$.

Proof. When all solutions of the difference equation of the linear terms approach zero all roots of the characteristic equation must be less than one in absolute value, hence $a_m \neq 0$. Some or all of the other a_i's may be zero. If $a_0 \neq 0$, this theorem is just a special case of Theorem 1: hence we assume that $a_0 = 0$. That is, we assume that the difference equation has the form

$$(12) \quad \begin{aligned} a_m S_{n+m} + a_{m-1} S_{n+m-1} + \cdots + a_k S_{n+k} + \phi_1(S_{n+m}, \cdots, S_n, n) \\ + a_{-1} n^{-1} = 0, \qquad k > 0. \end{aligned}$$

We replace S_{n+m} in Φ_1 by one of its values

$$S_{n+m} = f(S_{n+m-1}, \cdots, S_n, n),$$

which vanishes at the point $1/n = 0$, $S_{n+m-1} = 0, \cdots, S_n = 0$, and then set $u_1(n) = S_n, \cdots, u_m(n) = S_{n+m-1}$. This gives the system of equations

$$(13) \quad \begin{aligned} u_i(n+1) &= u_{i+1}(n) \qquad (i + 1, 2, \cdots, m-1) \\ u_m(n+1) &= -\frac{1}{a_m}[a_k u_{k+1}(n) + \cdots + a_{m-1} u_m \\ &\quad + \Phi_1(f, u_1(n), \cdots u_m(n), n) + a_{-1} n^{-1}] \end{aligned}$$

When the non-zero roots of the characteristic equation of the equation of the linear terms of (12) are distinct, we apply the linear transformation

$$(14) \quad \begin{aligned} u_i(n) &= z_i(n) \qquad (i = 1, 2, \cdots, k) \\ u_i(n) &= a_{i,k+1} z_{k+1}(n) + a_{i,k+2} z_{k+2}(n) + \cdots + a_{i,m} z_m(n) \\ &\qquad (i = k+1, \cdots, m) \end{aligned}$$

where the a_{ij} $(i, j = k + 1, \cdots, m)$ are so determined that the transformed system of difference equations has the form

$$(15) \quad \begin{aligned} z_i(n+1) &= z_{i+1}(n) \qquad (i = 1, 2, \cdots, k) \\ z_i(n+1) &= \rho_i z_i(n) + \Psi_i(z_1(n), \cdots, z_m(n), n) \qquad (i = k+1, \cdots, m). \end{aligned}$$

In order to clarify the proof we divide it into two parts. First we assume that the Ψ_i do not involve n explicitly. Under this assumption, for a given $\delta > 0$, however small, it is possible to choose h so small that

$$|\Psi_i(z_1(n), \cdots, z_m(n))| < \delta\{|z_1| + \cdots + |z_m|\} \qquad (i = k+1, \cdots, m)$$

when $|z_i| \leqq h$ $(i = 1, 2, \cdots, m)$. Therefore

$$|z_i(n+1)| < |\rho_i||z_i(n)| + \delta\{|z_1(n)| + \cdots + |z_m(n)|\} \leqq (|\rho_i| + m\delta)h \qquad (i = k+1, \cdots, m).$$

If $|\rho_j| \geqq |\rho_i|$, $(j \neq i)$, we set $|\rho_j| + m\delta = \sigma_1$ and choose δ so small that $\sigma_1 < 1$. Then,

$$|z_i(n+1)| < \sigma_1 h < h \qquad (i = k+1, \cdots, m).$$

It follows from the equations (15) that

$$|z_i(n+1)| \leqq h \qquad (i = 1, 2, \cdots, k),$$

and

$$|z_k(n+2)| < \sigma_1 h,$$

hence

$$|z_i(n+2)| < \sigma_1 h \qquad (i = k, \cdots, m)$$

$$|z_i(n+2)| < h \qquad (i = 1, 2, \cdots, k-1),$$

and so on until

$$|z_i(n+k+1)| < \sigma_1 h \qquad (i = 1, \cdots, m).$$

In general

$$|z_i(n+lk+1)| < \sigma_1^l h \qquad (i = 1, \cdots, m).$$

Hence the $z_i(n+lk)$ $(i = 1, \cdots, m)$ approach zero as l becomes infinite, consequently, there are sequences of (12) which approach zero as n becomes infinite.

Now if n appears explicitly in the Ψ_i, for a given $\delta > 0$, we may choose h so small and n_0 so large that

$$(16) \qquad |\Psi_i| < \delta(|z_1(n)| + \cdots + |z_m|) + M/n \qquad (i = k+1, \cdots, m)$$

for $|z_i| < h$ $(i = 1, \cdots, m)$ and $n \geqq n_0$. Hence for the initial conditions

$$|z_i(n_0)| < h,$$

$$|z_i(n_0+1)| < |\rho_i||z_i(n_0)| + \delta(|z_1(n_0)| + \cdots + |z_m(n_0)|) + M/n_0$$
$$< (|\rho_i| + m\delta)h + M/n_0 \qquad (i = k+1, \cdots, m).$$

If $|\rho_j| \geqq |\rho_i|$ for $i \neq j$, we set $|\rho_j| + m\delta = \sigma_1$ and choose δ so small that $\sigma_1 < 1$. After n_0 and h have been determined so that (16) is satisfied, we increase n_0 until $\sigma_1 + M/hn_0 = \sigma_2 < 1$. Then

$$|z_i(n_0+1)| < \sigma_1 h + M/n_0 < \sigma_2 h < h \qquad (i = k+1, \cdots, m)$$

$$|z_i(n_0+1)| < h \qquad (i = 1, \cdots, k)$$

and

$$|z_i(n_0+k+1)| < \sigma_1 h + M/n_0 < \sigma_2 h \qquad (i = 1, \cdots, m).$$

The z_i's may never exceed h in absolute value, and when l is so large that

$$\frac{M}{h\sigma_2(n_0 + l)} < \sigma_2 - \sigma_1$$

we have

$$|z_i(n_0 + 2k + l + 1)| < |\rho_i||z_i| + \delta(|z_i| + \cdots + |z_m|) + \frac{M}{n_0 + l} < \sigma_2^2 h \qquad (i = 1, \cdots, m).$$

Continuing step by step we see that the z_i $(i = 1, \cdots, m)$ are bounded by an expression which slowly approaches zero as n becomes infinite: hence they approach zero.

When the characteristic equation has equal roots, we employ a transformation (14) in which the a_{ij} $(i, j = k + 1, \cdots, m)$ are so determined that when $\rho_l = \rho_{l-1}$ the l^{th} equation of the transformed system is of the form

$$z_l(n + 1) = \rho_l z_l(n) + \epsilon z_{l-1}(n) + \Psi_l(z_1, z_2, \cdots, z_m, n), \tag{17}$$

where ϵ is chosen so small that $|\rho_j| + \epsilon < 1$, $|\rho_j| \geqq |\rho_i|$, $i \neq j$, and the remaining equations are of the form (11). Treating this new system by a method analogous to the one used above, in which $|\rho_j| + \epsilon$ plays the role corresponding to that of $|\rho_j|$, we see that it has solutions which approach zero. Hence our proof is complete.

Theorems 1 and 2 exclude difference equations in which the degree of n in the coefficient of the non-linear terms is higher than the degree of n in the coefficients of the linear terms. In regard to such equations we have the following theorem.

THEOREM 3. *An algebraic difference equation, whose terms are of at least first degree in the S_{n+i}, defines sequences which approach zero as n becomes infinite, if all roots of the characteristic equation of the equation of the linear terms are less than one in absolute value*

PROOF. As in the proof of Theorems 1 and 2, we replace S_{n+m} by one of its values that vanishes at the point $S_n = 0, \cdots, S_{n+m-1} = 0$, and then we apply a linear transformation which transforms this new equation into a system of equations of the form (11), (15) or (17).

CASE I. The coefficients of the linear terms in the S_{n+i} are constants.

When all roots of the characteristic equation are distinct and different from zero, the system has the form

$$z_i(n + 1) = \rho_i z_i(n) + \Psi_i(z_1(n), \cdots, z_m(n), n) \qquad (i = 1, \cdots, m),$$

where all terms of the Ψ_i $(i = 1, \cdots, m)$ are of at least second degree in the z_i. Hence there exists an h such that

$$|\Psi_i(z_1, \cdots, z_m, n)| < Mn^l(|z_1| + \cdots + |z_m|)^2 \qquad (i = 1, \cdots, m)$$

when $|z_i| < h$ $(i = 1, \cdots, m)$, where M and l are constants. If $|\rho_j| \geqq |\rho_i|$

for all i, we choose a number σ such that $|\rho_i| < \sigma < 1$, and then choose n_0 so large that

$$|m^2 M n_0^l \sigma^{n_0}| + |\rho_j| \leqq \sigma.$$

Such a choice of n_0 is possible since $\lim_{n\to\infty} n^l \alpha^n = 0$, if $|\alpha| < 1$. Now for z_i in the region $|z_i| \leqq h \leqq \sigma^{n_0}$ $(i = 1, \cdots, m)$

$$\begin{aligned} |z_i(n_0+1)| &\leqq |\rho_i||z_i| + M n_0^l m^2 \sigma^{2n_0} \\ &\leqq (|\rho_i| + M n_0^l m^2 \sigma^{n_0})\sigma^{n_0} \leqq \sigma^{n_0+1} \qquad (i = 1, 2, \cdots, m) \\ |z_i(n_0+2)| &\leqq |\rho_i||z_i(n_0+1)| + M(n_0+1)^l m^2 \sigma^{2(n_0+1)} \\ &\leqq (|\rho_i| + M(n_0+1)^l m^2 \sigma^{n_0+1})\sigma^{n_0+1} \\ &< (|\rho_i| + M n_0^l m^2 \sigma^{n_0})\sigma^{n_0+1} < \sigma^{n_0+2} \qquad (i = 1, \cdots, m), \end{aligned}$$

and in general

$$|z_i(n_0+p)| < \sigma^{n_0+p} \qquad (i = 1, 2, \cdots, m).$$

Hence the $z_i(n_0+p)$ approach zero as p becomes infinite, and consequently $\{S_n\} \to 0$ as $n \to \infty$.

When some of the ρ_i equal zero the transformed system of equations is of the form (15). Again we choose σ such that $|\rho_j| < \sigma < 1$, but now we take n_0 so large that

$$m^2 M(\mu n_0)^l \sigma^{n_0} + |\rho_j| \leqq \sigma, \text{ where } \mu \text{ is the maximum of } k \text{ and } 2.$$

Then for $|z_i(n_0)| \leqq \sigma^{n_0}$ $(i = 1, \cdots, m)$

$$\begin{aligned} |z_i(n_0+1)| &\leqq |\rho_i||z_i| + M n_0^l m^2 \sigma^{2n_0} \\ &\leqq (|\rho_j| + M n_0^l m^2 \sigma^{n_0})\sigma^{n_0} \leqq \sigma^{n_0+1} \qquad (i = k+1, \cdots, m) \\ |z_i(n_0+1)| &\leqq \sigma^{n_0} \qquad (i = 1, 2, \cdots, k): \\ |z_i(n_0+2)| &\leqq |\rho_j||z_i| + M(n_0+1)^l m^2 \sigma^{2n_0+2} \\ &\leqq \sigma^{n_0}(|\rho_j| + M m^2 (\mu n_0)^l \sigma^{n_0}) \leqq \sigma^{n_0+1} \qquad (i = k, \cdots, m) \\ |z_i(n_0+2)| &\leqq \sigma^{n_0} \qquad (i = 1, 2, \cdots, k-1): \end{aligned}$$

and

$$\begin{aligned} |z_i(n_0+k+1)| &\leqq |\rho_j||z_i| + M(n_0+k)^l m^2 \sigma^{2(n_0+k)} \\ &\leqq (|\rho_j| + M n^2 (\mu n_0)^l \sigma^{n_0})\sigma^{n_0} \leqq \sigma^{n_0+1} \qquad (i = 1, \cdots, m). \end{aligned}$$

Consequently,

$$|z_i(n_0+k+2)| \leqq |\rho_j||z_i| + M(n_0+k+1)^l m^2 \sigma^{2(n_0+k+1)} \leqq \sigma^{n_0+2}$$
$$(i = k+1, \cdots, m)$$

$$| z_i(n_0 + 2k + 1) | \leqq \sigma^{n_0+2} \qquad (i = 1, \cdots, m).$$

Continuing step by step we have

$$| z_i(n_0 + pk + 1) | \leqq \sigma^{n_0+p} \qquad (i = 1, \cdots, m).$$

Hence the $z_i(n_0 + pk + 1)$ approach zero as $p \to \infty$, and the corresponding sequence $\{S_n\} \to 0$ as $n \to \infty$.

As in the proof of Theorem 2, the above treatment, with suitable modifications, applies to the case where the characteristic equation has multiple roots.

CASE II. The coefficients of the linear terms in S_{n+i} are polynomials in n. The transformed system of equations may be written in the form

$$z_i(n + 1) = z_{i+1}(n) \qquad (i = 1, \cdots, k)$$

$$(18) \quad z_i(n + 1) = \rho_i z_i(n) + b_{i1}(1/n)z_1(n) + \cdots + b_{im}(1/n)z_m(n) + \Psi_i(z_1(n), \cdots, z_m(n), n) \qquad (i = k + 1, \cdots, m),$$

where the b_{ij} are convergent series in $1/n$, whose constant terms are zero, and all terms of the Ψ_i $(i = k + 1, \cdots, m)$ are of at least second degree in the arguments z_i $(i = 1, \cdots, m)$. We first choose n_1 so large that

$$| \rho_j | + mb = \rho < 1$$

where $b \geqq | b_{ij}(1/n) |$ $(i, j = 1, \cdots, m)$ for $n > n_1$. Then the argument in case I holds where $| \rho_j |$ is replaced by ρ and n_0 is chosen greater than n_1. Hence our proof is complete.

2. **Sequences with finite limits.** The only possible finite limits for the sequences defined by an algebraic difference equation, with constant coefficients,

$$(19') \qquad \Phi(S_{n+m}, S_{n+m-1}, \cdots, S_n) = 0$$

are the roots of the algebraic equation

$$(20') \qquad \Phi(S, S, \cdots, S) = 0,$$

where the latter is obtained from the former by replacing S_{n+i} $(i = 1, \cdots, m)$ by S. Likewise, the only possible finite limits for the sequences defined by an algebraic difference equation

$$(19) \qquad \Phi_1(S_{n+m}, \cdots, S_n, n) = 0$$

are the roots of the equation

$$(20) \qquad \Phi_1'(S, \cdots, S) = 0,$$

where $\Phi_1' = \lim_{n \to \infty} n^{-q} \Phi_1(S_{n+m}, \cdots, S_n, n)$ and q is the highest power of n appearing in Φ_1.

If $S = \alpha$ is a root of the equation (20), and if the difference equation

$$(21) \qquad \bar{\Phi}(\bar{S}_{n+m}, \cdots, \bar{S}_n, n) = 0,$$

obtained from (19) by the transformation $S_n = \bar{S}_n + \alpha$, satisfies the hypothesis of theorems 1, 2 or 3, then (19) defines sequences which approach α. For the sake of brevity, we shall call the difference equation (21) the "*transformed equation.*"

THEOREM 4. *If α is an algebraic number, then there exist algebraic difference equations that define sequences which approach α.*[6]

To construct one: it is sufficient to take an algebraic difference equation, involving arbitrary parameters in such a manner that the equation (20) has a root α, and then to choose values for the parameters so that there are sequences defined by the transformed equation (21) which approach zero.

To illustrate the method, we construct a difference equation of second order that defines sequences which approach $\beta^{\frac{1}{2}}$, where β is a rational number. The difference equation

$$aS_nS_{n+1} + bS_{n+1}S_{n+2} - cS_{n+1}^2 - (b + a - c)\beta = 0 \tag{22}$$

may define sequences which approach $\beta^{\frac{1}{2}}$ or $-\beta^{\frac{1}{2}}$, since its limit as $n \to \infty$ is $S^2 = \beta$. Applying the transformation $S_n = \bar{S}_n + \beta$, we obtain for the equation of the linear terms of the transformed equation

$$b\bar{S}_{n+2} + (b + a - 2c)\bar{S}_{n+1} + a\bar{S}_n = 0. \tag{23}$$

If we choose a, b and c such that the roots of the characteristic equation of (23) are less than one in absolute value, there are sequences of (22) which converge to $\beta^{\frac{1}{2}}$. For example when $a = 1$, $b = 6$, $c = 6$, the roots of the characteristic equation (23) are 1/2 and 1/3, hence some sequences of (21) will converge to $\beta^{\frac{1}{2}}$. In particular, the difference equation

$$S_nS_{n+1} + 6S_{n+1}S_{n+2} - 6S_{n+1}^2 - 2 = 0 \tag{24}$$

defines sequences which approach $\sqrt{2}$. For the initial values $S_0 = 1$, $S_1 = 2$ and $S_0 = 1.4$, $S_1 = 1.4$ the sequences are

$$1, 2, 7/6, 9/7, 1021/756, \cdots ;$$

and

$$1.4,\ 1.4,\ 59/42 = 1.404+,\ 34908/24780$$
$$= 1.4087+,\ 4804998967/3395410200 = 1.4122+, \cdots$$

respectively. If instead of the above values we take $a = 1$, $b = 54$, $c = 35$, $\beta = 2$, we obtain the difference equation

$$54S_{n+1}S_{n+2} - 35S_{n+1}^2 + S_{n+1}S_n - 40 = 0. \tag{25}$$

[6] The truth of this theorem follows immediately from the work of Schröder (4) on the approximation of the roots of an algebraic equation by the iteration of rational functions.

For the initial conditions $S_0 = 1$, $S_1 = 2$ and $S_0 = 1.4$, $S_1 = 1.4$ the sequences are

$$1, 2, 23/38, 30647/22356 = 1.326+, 1.4052+, \cdots$$

and

$$1.4, 1.4, 1333/985 = 1.4105+, 48074279/34011495 = 1.4134+, \cdots$$

respectively.

The sequences defined by (25) converge more rapidly to the $\sqrt{2}$ than the corresponding sequences defined by (24). The question arises, is there a difference equation whose sequences approach a given algebraic number more rapidly than those defined by all other difference equations? Under the conditions for which the generalized Poincare theorem[3] holds, the smaller the absolute values of the roots—all greater than zero—of the characteristic equation of the equation of the linear terms the more rapid the convergence to zero of the solution of the difference equation. Hence we were led to predict the following theorem.

THEOREM 5. *Of the family of non-linear difference equations of the form*

$$a_m(n)S_{n+m} + a_{m-1}(n)S_{n+m-1} + \cdots + a_0(n)S_n + \Phi(S_{n+m}, \cdots, S_n, n) = \theta(n); \tag{26}$$

where Φ is a fixed polynomial whose terms are of at least second degree in the S_{n+i}, the $a_i(n)$ $(i = 0, 1, \cdots, m)$ and $\theta(n)$ are polynomials in n, $a_m(n) \not\equiv 0$, and the coefficients of $a_m(n)$ are fixed while those of the $a_i(n)$ $(i = 0, 1, \cdots, m - 1)$ and $\theta(n)$ are arbitrary; the one that defines sequences which converge most rapidly to zero, for arbitrary initial conditions from a sufficiently small region surrounding the origin, is the one in which the $a_i(n)$, $(i = 0, 1, \cdots, m - 1)$ and $\theta(n)$ are identically equal to zero.

PROOF. (a) First, $\theta(n)$ must vanish. When $\theta(n) \not\equiv 0$ the sequences cannot approach zero more rapidly than n^{-k}, where k is an integer. For if this were true, the left member of (26) would approach zero more rapidly than the right member which is impossible. The proof of Theorem 3 reveals that there are equations of form (26) which define sequences that approach zero as rapidly as $Mn^l\sigma^n$. Therefore if the sequences have the maximum rate of convergence, $\theta(n) \equiv 0$.

(b) Second, the degree of n in $a_m(n)$ must be greater than or equal to the degree of n in all the other a's. Otherwise, there is no region of strong convergence.

(c) Third, if the coefficients of the terms of highest degree in n of $a_0(n), \cdots, a_m(n)$ are $a_{00}, a_{10}, \cdots, a_{m0}$, respectively, then the difference equation of the family whose sequences converge most rapidly to zero is the one for which $a_{m0} \neq 0$ and $a_{i0} = 0$, $(i = 0, 1, \cdots, m - 1)$.

To prove this we refer back to the notation and the statements in the proof of Theorem 3. When there is one a_{i0} that is different from zero there is at least one ρ_i that is not equal to zero. It is only necessary to consider the case where all of the ρ_i are less than one in absolute value. For it is possible to show, by

a slight alteration of the inequalities in the proof of Theorem 3, that of the sequences defined by two difference equations, identical except for the ρ_i of maximum absolute value, those defined by the difference equation with the smaller ρ_i converge more rapidly than *some* of those defined by the other equation.

So we assume that $1 > |\rho_j| > |\rho_i|$ $(i \neq j)$, and consider the equations (18). We choose n_0 so large that

$$\begin{aligned} m^2 M(\mu n)^l \sigma^n &< \tfrac{1}{2}(|\rho_j|/3)^{m+1} \\ &< \frac{|\rho_j| - |\rho_i|}{2} \qquad (|\rho_l| \neq |\rho_r|;\ l, r = 1, \cdots, m) \\ &< \epsilon/2 \qquad \text{when } \rho_l = \rho_{l-1}, \end{aligned}$$

where ϵ is the ϵ of equation (17);

$$m\,|\,b_{ij}(n)\,| \leqq mb \leqq \tfrac{1}{2}(|\rho_j|/3)^{m+1};$$

$$\rho_j + mb = \rho < 1;$$

and

$$m^2 M(\mu n)^l \sigma^n + \rho < \sigma < 1, \qquad n \geqq n_0 .$$

When the initial conditions are $|\,z_i(n_0)\,| \leqq \sigma^{n_0}$, the last two conditions are sufficient to insure that the sequences approach zero such that

$$|\,z_i(n_0 + pk + 1)\,| < \sigma^{n_0+p}.$$

When $z_j(n_0) = \sigma^{n_0}$ and $|\,z_i(n_0)\,| \leqq \sigma^{n_0}$, $(i = 1, 2, \cdots, m)$ we have in succession:

$$\begin{aligned} |\,z_j(n_0+1)\,| &\geqq |\,z_j(n_0)\,|\,|\,\rho_j\,| - b\sum_{r=1}^{m-1} |\,z_r(n_0)\,| - M(n_0)^l \left[\sum_{r=1}^{m-1} |\,z_r(n_0)\,|\right]^2 \\ &\geqq |\,z_j(n_0)\,|\,[\,|\,\rho_j\,| - (|\,\rho_j\,|/3)^{m+1}] \geqq z_j(n_0)(2|\,\rho_j\,|/3), \\ |\,z_j(n_0+2)\,| &\geqq |\,z_j(n_0+1)\,|\,|\,\rho_j\,| \\ &\qquad - b\sum |\,z_r(n_0+1)\,| - M(n_0+1)^l[\sum |\,z_r(n_0+1)\,|]^2 \\ &\geqq |\,z_j(n_0+1)\,|\left[\,|\,\rho_j\,| - \frac{\max |\,z_r(n_0+1)\,|}{|\,z_r(n_0+1)\,|}\cdot\left(\frac{|\,\rho_j\,|}{3}\right)^{m+1}\right] \\ &\geqq |\,z_j(n_0+1)\,|\left[\,|\,\rho_j\,| - \left(\frac{|\,\rho_j\,|}{3}\right)^{m}\right] \geqq |\,z_j(n_0+1)\,|\cdot\left(\frac{2|\,\rho_j\,|}{3}\right), \end{aligned}$$

. . . .

$$|\,z_j(n_0+l)\,| \geqq |\,z_j(n_0)\,|\left(\frac{2|\,\rho_j\,|}{3}\right)^l, \qquad (l = 2, 3, \cdots, m+1)$$

Under the above restrictions,

$$\frac{|\,z_i(n)\,|}{|\,z_j(n)\,|} \leqq \left(\frac{2|\,\rho_j\,|}{3}\right)^{-m}, \qquad n \geqq n_0$$

Thus,

$$
\begin{aligned}
&|z_j(n_0+m+2)| \\
&\geqq |z_j(n_0+m+1)| \Big[|\rho_j| - \{b - M(n_0+m+1)^l \sum |z_r(n_0+m+1)|\} \\
&\qquad\qquad \cdot \frac{\sum |z_r(n_0+m+1)|}{|z_j(n_0+m+1)|} \Big] \\
&\geqq |z_j(n_0+m+1)| \left[|\rho_j| - (|\rho_j|/3)^{m+1} \left(\frac{2|\rho_j|}{3}\right)^{-m} \right] \\
&\geqq |z_j(n_0+m+1)| \left(\frac{2|\rho_j|}{3}\right) \\
&\geqq z_j(n_0)(2|\rho_j|/3)^{m+2}.
\end{aligned}
$$

Continuing, we obtain

$$|z_j(n_0+m+l)| \geqq z_j(n_0)(2|\rho_j|/3)^{l+m} \qquad l = 0, 1, \cdots .$$

Therefore, if some ρ_i is different from zero, *some* sequences defined by the difference equation converge to zero more slowly than the sequence $(2|\rho_j|/3)^n$. While if $\rho_i = 0$, $i = 1, 2, \cdots, m$, all sequences, with initial conditions $|z_i(n_0)| \leqq \sigma^{n_0}$, converge to zero more rapidly than the sequence $(|\rho_j|/3)$. Thus for the most rapid convergence of all sequences, with arbitrary initial conditions, $a_{i0} = 0$ $(i = 0, 1, \cdots, m-1)$.

(d) Fourth, the $a_i(1/n) \equiv 0$ $(i = 0, 1, \cdots, m-1)$. For assume this statement is false and set $S_n = n^{\lambda n} T_n$, where λ is so determined that in the new equation one other linear term is of the same degree in n as $a_m(n)$. The quantity λ will be negative. Remove the common factor $n^{\lambda n}$. Now, although this new equation is not an algebraic equation, the inequalities are stronger than they were for the proof of (c) above. Therefore the argument of (c) applies. Hence by a repetition of the argument if the sequences have the maximum rate of convergence $a_i(1/n) \equiv 0$ $(i = 0, 1, \cdots, m-1)$. Q.E.D.

Let us again consider the example. According to Theorem 5 there is a difference equation of the form (22) that defines sequences which converge more rapidly to $\sqrt{2}$ than those defined by (25). It is the one for which $a = 0$, $b + a - 2c = 0$. That is, the equation

$$2S_{n+1}S_{n+2} - S_{n+1}^2 - 2 = 0,$$

or

$$2S_nS_{n+1} - S_n^2 - 2 = 0. \tag{27}$$

For the initial values $S_0 = 1$ and $S_0 = 1.4$ the sequences are

$$1, 1.5, 1.4166+, 577/408 = 1.41421\overset{*}{5}+, 665857/470832 = 1.41421356237\overset{*}{6}+ \cdots$$

$$886731088897/627013566048 = 1.4142135623730950 \cdots +, \cdots$$

an

$$7/5 = 1.4,\ 99/70 = 1.4142\overset{*}{8}+,\ 19601/13860 = 1.41421356\overset{*}{4},$$

$$768398401/543338720 = 1.41421356237309504\overset{*}{9}9+$$

respectively. The asterisk indicates the first incorrect decimal of the terms. The sequences defined by this equation converge to the limit very rapidly. If we assume the error of the n^{th} term is ϵ and set $S_n = 2^{\frac{1}{2}} + \epsilon$, then

$$S_{n+1} = 2^{\frac{1}{2}} + 2^{-\frac{3}{2}}\epsilon^2 - 2^{-2}\epsilon^3 + \cdots .$$

Hence, when S_n is sufficiently close to the limit the error is approximately squared with each successive term of the sequence. We say that the order of rate of convergence of the sequences is 2.

The difference equation (27) is of first order, and we know from the work of Schröder[4] on iteration of rational functions that it is possible to set up a first order difference equation with constant coefficients, whose sequences approach a given algebraic number α with an *order of rate of convergence* k, where k is any integer. That is, there exists a first order difference equation with the property that when S_n is replaced by $\alpha + \epsilon$, then

$$S_{n+1} = \alpha + c_1\epsilon^k + c_2\epsilon^{k+1} + \cdots ,$$

where c_1 is a constant independent of ϵ. (If $k = 1$, $|c_1| < 1$).

Is this true for difference equations of higher than first order?

THEOREM 6. *For a given integer k and an algebraic number α, there exist algebraic difference equations of the m^{th} order that define rational sequences which approach α with an order of rate of convergence k.*

PROOF. Let

$$S_{n+m} = \Phi(S_n, \cdots, S_{n+m-1}, n) \tag{28}$$

be an algebraic difference equation, where Φ, a rational function in its arguments $S_n, \cdots, S_{n+m-1}, n$, has the Taylor's expansion

$$\begin{aligned}\Phi = a_0 + a_{10}(S_n - \alpha) + a_{11}(S_{n+1} - \alpha) + \cdots + a_{1m-1}(S_{n+m-1} - \alpha)\\ + a_{200}(S_n - \alpha)^2 + a_{210}(S_{n+1} - \alpha)(S_n - \alpha) + \cdots\end{aligned} \tag{29}$$

which converges in some neighborhood of the point $S_n = \alpha, \cdots, S_{n+m-1} = \alpha$, $n = \infty$. (The a's are rational functions of n.)

There are sequences defined by (28) which converge to α with an order of rate of convergence k, if, when $S_n = \alpha + \epsilon$, $S_{n+1} = \alpha + \theta_1\epsilon^k, \cdots, S_{n+m-1} = \alpha + \theta_{m-1}\epsilon^{k^{m-1}}$, where the θ_i $(i = 1, \cdots, m - 1)$ are constants independent of ϵ, then $S_{n+m} = \alpha + \theta_m\epsilon^{k^m} + c_2\epsilon^{k^m+1} + \cdots$. This is true, when $a_0 = \alpha$ and $a_{li_1,\cdots,i_l} \equiv 0$ for all values of l for which

$$k^{i_1} + k^{i_2} + \cdots + k^{i_l} < k^m \qquad (i_1, i_2, \cdots, i_l < m) \tag{30}$$

and when $a_{l i_1, \cdots, i_l} \neq 0$ for some values of $i_1, \cdots, i_l$ such that $k^{i_1} + k^{i_2} + \cdots + k^{i_l} = k^m$. Hence, if there exists a rational function Φ such that $\Phi(\alpha, \cdots, \alpha, n) \equiv \alpha$ and

$$\left. \frac{\partial^{j_1+j_2+\cdots+j_m} \Phi}{\partial S_n^{j_1} \partial S_{n+1}^{j_2} \cdots \partial S_{n+m-1}^{j_m}} \right|_{\alpha} \equiv 0$$

for all j's such that

$$j_1 + kj_2 + k^2 j_3 + \cdots + k^{m-1} j_m < k^m$$

then the theorem is true.

We state that if α is a simple root of an algebraic equation $f(z) = 0$, then

$$\begin{aligned} &\Phi(S_n, S_{n+1}, \cdots, S_{n+m-1}, n) \\ &= \frac{1}{m}\left[S_n + \sum_{j=1}^{k^m-1} \left\{(-1)^j \frac{f^j(S_n)}{j!} \left(\frac{1}{f'(S_n)} \frac{d}{dS_n}\right)^{j-1} \frac{1}{f'(S_n)}\right\} + f^{k^m}(S_n)\phi_0\left(S_n, \frac{1}{n}\right)\right] \\ &\quad + \frac{1}{m}\left[S_{n+1} + \sum_{j=1}^{k^{m-1}-1} \left\{(-1)^j \frac{f^j(S_{n+1})}{j!} \left(\frac{1}{f'(S_{n+1})} \frac{d}{dS_{n+1}}\right)^{j-1} \frac{1}{f'(S_{n+1})}\right\} \right. \\ &\qquad\qquad \left. + f^{k^{m-1}}(S_{n+1})\phi_1\left(S_{n+1}, \frac{1}{n}\right)\right] \\ &\quad + \cdots + \frac{1}{m}\left[S_{n+m-1} + \sum_{j=1}^{k-1} \left\{(-1)^j \frac{f^j(S_{n+m-1})}{j!} \left(\frac{1}{f'(S_{n+m-1})} \frac{d}{dS_{n+m-1}}\right)^{j-1} \right.\right. \\ &\qquad\qquad \left.\left. \cdot \frac{1}{f'(S_{n+m-1})}\right\} + f^k(S_{n+m-1})\phi_{m-1}\left(S_{n+m-1}, \frac{1}{n}\right)\right] \end{aligned}$$

is a function[7] satisfying the above conditions, where the ϕ_i $(i = 0, 1, \cdots, m-1)$ are arbitrary polynomials which are finite at the point $S_{n+i} = \alpha$ $(i = 1, \cdots, m-1)$ and $n = \infty$, f' is the first derivative of f with respect to its variable, f^j is the j^{th} power of f, and $((1/f'(z))\, d/dz)^j$ is a symbolic operator which subjects $1/f'(z)$ to the operation of differentiation and multiplication j times. For example

$$\left(\frac{1}{f'} \frac{d}{dz}\right)^4 \frac{1}{f'} = \frac{1}{f'} \frac{d}{dz}\left(\frac{1}{f'} \frac{d}{dz}\left\{\frac{1}{f'} \frac{d}{dz}\left[\frac{1}{f'} \frac{d}{dz}\left(\frac{1}{f'}\right)\right]\right\}\right).$$

First, $\Phi(\alpha, \alpha, \cdots, \alpha, n) = \alpha$, for when $S_{n+i} = \alpha$, $(i = 1, \cdots, m-1)$, all terms, except the first in each of the brackets, contain a factor $f(\alpha)$ and hence vanish. Second, the $k^m - 1$ partial derivatives with respect to S_n vanish at the point $S_{n+i} = \alpha$, $(i = 1, \cdots, m-1)$, since the derivatives of the expressions in the last $m - 1$ brackets are zero and Schröder has proved that the $k^m - 1$ derivatives of the expression in the first bracket vanish when $S_n = \alpha$. Similarly, the $k^{m-i} - 1$ partial derivatives of Φ with respect to S_{n+i} vanish when

[7] This is a generalization of a formula developed by Schröder.

$S_{n+1} = \alpha$. And lastly, all second partial derivatives with respect to two different S_{n+i} vanish identically. Q.E.D.

ASSUMPTION. In order to assure that the sequences under consideration are rational we shall assume, in the future, that S_{n+m} enters the equation only in the first degree.

3. The effect of the order of a difference equation and the degree of n in the coefficients upon the order of rate of convergence of the sequences. It follows from the work of Schröder, and from the above theorem, that it is possible to increase the order of rate of convergence of sequences, defined by difference equations, by increasing the degree of the equations. Now we ask the questions: When the degree is held constant, how does a change in the order of the difference equation affect the order of the convergence of the sequences? What is the relation between the degree of n in the coefficients of a difference equation and the order of convergence of the sequences?

As a partial answer to the last question we have the following:

THEOREM 7. *The order of rate of convergence, r, of rational sequences defined by a first order difference equation whose coefficients are polynomials in n, cannot exceed the order of convergence of rational sequences defined by* SOME *first order difference equation of the same degree with constant coefficients, provided $r \geqq 2$.*

PROOF. Again without loss in generality, we consider only sequences which approach zero. Thus reducing our problem to that of showing that, "if S_n is a null sequence defined by a difference equation of order one and degree q, then the order of rate of convergence r of S_n is $<q + 1$ and it is equal to q for some equations of this type which have constant coefficients."

When the coefficients are constants, the maximum rate of convergence of the sequences defined by first order difference equations of the q^{th} degree is q. If the sequences do converge to zero with this maximum rate of convergence then the Taylor's expansion of the left member of the equation

$$S_{n+1} = \Phi(S_n),$$

gives

$$S_{n+1} = a_q S_n^q + a_{q+1} S_n^{q+1} + \cdots$$

where $a_q \neq 0$. If the coefficients of Φ were polynomials in n then the a_i would be rational functions of n. In order for the sequences of such an equation to have the maximum rate of convergence $a_i \equiv 0$, $(i = 1, \cdots, q - 1)$; $a_q(n) \not\equiv 0$, and $a_q(n)$ must be a rational function of n in which the degree of the numerator is less than the degree of the denominator, viz.

$$a_q(n) = a_\lambda n^{-\lambda} + a_{\lambda+1} n^{-\lambda-1} + \cdots \qquad (\lambda > 0),$$

and the equation may be written in the form

$$(31) \quad S_{n+1} = n^{-\lambda} S_n^q [(a_\lambda + a_{\lambda+1} n^{-1} + \cdots) + n^\lambda a_{q+1}(n) S_n + n^\lambda a_{q+2}(n) S_n^q + \cdots].$$

Hence if the order of rate of convergence of the sequences may be increased by allowing the coefficients to be polynomials in n, the order of convergence of the sequences of (31) must be greater than that of the sequences defined by the equation

$$S_{n+1} = a_\lambda S_n^q .$$

That is, the order of rate of convergence must equal or exceed the order of convergence of the sequences of

$$T_{n+1} = a_\lambda T_n^{q+1}. \tag{32}$$

Now when the sequences converge with an order of rate $\geqq 2$, we may choose n_0 so large and then S_{n_0} so small that

$$a_\lambda^{-1} \mid a_{\lambda+1} n^{-1} + a_{\lambda+2} n^{-2} + \cdots \mid < \epsilon/2 < \tfrac{1}{2}, \qquad n > n_0 ,$$

$$n^\lambda a_\lambda^{-1} \mid a_{q+1}(n) S_n + a_{q+2}(n) S_n^2 + \cdots \mid < \epsilon/2 < \tfrac{1}{2}, \qquad n > n_0 ,$$

with such a choice the sequences defined by (31) are term by term greater in absolute value than the sequences defined by the difference equation

$$\bar{S}_{n+1} = n^{-\lambda} a_\lambda (1 - \epsilon) \bar{S}_n^q . \tag{31'}$$

The solution of this equation for $\bar{S}_{n_0} = S_{n_0}$ is

$$\bar{S}_{n_0+p} = S_{n_0}^{q^p} a_\lambda^{(q^p-1)/q} (1 - \epsilon)^{(q^p-1)/q} (p + n_0 - 1)^{-\lambda} (p + n_0 - 2)^{-\lambda q} \cdots (n_0)^{-\lambda q^{p-1}}$$

and the solution of (32) is

$$T_{n_0+p} = a_\lambda^{((q+1)^p-1)/q} S_{n_0}^{(q+1)^p}.$$

But

$$\lim_{p\to\infty} \left(\frac{T_{n_0+p}}{\bar{S}_{n_0+p}} \right) = \lim_{p\to\infty} (S_{n_0})^{(q+1)^p - q^p} a_\lambda^{((q+1)^p - q^p)/q} (1 - \epsilon)^{(q^p-1)/q} (p + n_0 - 1)^\lambda$$
$$\cdot (p + n_0 + 2)^{\lambda q} \cdots (n_0)^{\lambda q^{p-1}} = 0.$$

Therefore the sequences of (31′), and consequently those of (31), do not converge to zero as rapidly as those of (32), so they cannot have an order of rate of convergence $q + 1$. Q.E.D.

Although we have been unable to generalize Theorem 7 we should like to conjecture that it holds for m^{th} order equations.

Now let us concentrate upon the relation between the order of difference equations and the order of the rate of convergence of their sequences.

THEOREM 8. *There exist algebraic difference equations of higher than first order, that define rational sequences which converge to certain algebraic numbers with a higher order of convergence than the sequences defined by any first order difference equations of the same degree.*

PROOF. Take the general equation of the m^{th} order and q^{th} degree with constant coefficients,

$$\sum_{j=1}^{q} \sum_{i_1=0}^{m-1} \sum_{i_2 \geqq i_1}^{m-1} \cdots \sum_{i_j \geqq i_{j-1}}^{m} (c_{0 i_1 i_2 \cdots i_j} S_{n+i_1} S_{n+i_2} \cdots S_{n+i_j}) + c_0 = 0,$$

or

$$(33)\qquad S_{n+m} = -\frac{\sum_{j=1}^{q}\sum_{i_1=0}^{m-1}\sum_{i_2\geq i_1}^{m-1}\cdots\sum_{i_j\geq i_{j-1}}^{m-1}(c_{0i_1i_2\cdots i_j}S_{n+i_1}S_{n+i_2}\cdots S_{n+i_j}) + c_0}{\sum_{j=1}^{q}\sum_{i_1=0}^{m-1}\sum_{i_2\geq i_1}^{m-1}\cdots\sum_{i_{j-1}\geq i_{j-2}}^{m-1}(c_{0i_1i_2\cdots i_j}S_{n+i_1}S_{n+i_2}\cdots S_{n+i_j}) + c_{0m}}$$

and demand that it define sequences which approach the number $\beta^{1/q}$ with an order of rate of convergence k, where k is an integer to be determined, β is a rational number, and $\beta^{j/q}$ is irrational $(j = 1, \cdots, q - 1)$. That is, set $S_{n+j} = \beta + \theta_j\epsilon^{k^j}$ $(j = 0, 1, \cdots, m - 1)$, expand the right member of (33) in powers of ϵ and demand that the constant term equal $\beta^{1/q}$, and that the coefficients of ϵ^t vanish for $t = 1, 2, \cdots, k^m - 1$. These demands impose k^m linear homogeneous relations upon the coefficients of the given equation (33); as can readily be seen by dividing the numerator by the denominator. The k^m linear relations may contain some of the θ_i $(i = 0, \cdots, m - 1)$ and some fractional powers of β. In order to insure that the coefficients of the difference equation (33) be rational, the coefficients of θ_i and the fractional power of β in these linear relations must vanish.

The maximum value for k is determined by the number of the relations that the coefficients of (33) may satisfy, that is, by the number of coefficients. This number[8] is

$$\sum_{j=1}^{q}\frac{m(m+1)\cdots(m+j-2)(m+2j-1)}{j!} + 1.$$

Since, when $m = 1$ it is $\sum_{j=1}^{q}[2] + 1 = 2q + 1$, the coefficients of a first order difference equation may be chosen so as to satisfy at least $2q$ linear homogeneous relations. Therefore k may be at least 2, for when $k = 2$ the coefficients are only restricted by $2q$ relations, q relations which insure that the constant term in the power series in ϵ equal $\beta^{1/q}$ and q relations which insure that the coefficient of ϵ equal zero. Hence if a difference equation of higher than first order defines sequences whose order of rate of convergence, k, is greater than that of sequences defined by a first order difference equation, k must be at least 3.

---

[8] In order to see this, write

$$\sum_{i_1=0}^{m-1}\cdots\sum_{i_j\geq i_{j-1}}^{m} = \sum_{i_1=0}^{m-1}\cdots\sum_{i_j\geq i_{j-1}}^{m-1} + \sum_{i_1=0}^{m-1}\cdots\sum_{i_{j-1}\geq i_{j-2}}^{m-1}.$$

The number of terms in the first sum on the right is $m(m+1)\cdots(m+j-1)/j!$. Therefore, the total number of terms of j^{th} degree is

$$\frac{m(m+1)\cdots(m+j-1)}{j!} + \frac{m(m+1)\cdots(m+j-2)}{(j-1)!} = \frac{m(m+1)\cdots(m+j-2)(m+2j-1)}{j!}.$$

Now consider difference equation of the second order. When $m = 2$ the number of terms in (33) is

$$\sum_{j=1}^{q} (2j + 1) + 1 = \frac{2q(q + 1)}{2} + q + 1 = (q + 1)^2.$$

If a sequence converges to $\beta^{1/q}$ with an order of rate of convergence 3, the coefficients must satisfy $18q$ relations—q relations to make the first term of the series in ϵ equal $\beta^{1/q}$, $2q$ to make the coefficients of ϵ and ϵ^2 vanish; $6q$ to make the coefficients of ϵ^3, ϵ^4, ϵ^5 vanish (twice as many here since θ_1 may be irrational and may appear in each of these relations); and $9q$ to make the coefficients of ϵ^6, ϵ^7 and ϵ^8 vanish (θ_1 and θ_1^2 may enter in each of these three relations). Since there are $(q + 1)^2$ coefficients in the difference equation, if $q \geqq 17$ these $18q$ relations may be satisfied, for then the number of homogeneous linear equations is less than the number of undetermined coefficients. Hence, for $q \geqq 17$ there exists a second order difference equation which defines sequences that approach $\beta^{1/q}$ with an order of rate of convergence 3.

Now if we can prove that no first order difference equation of 17th degree may define sequences which approach $\beta^{1/17}$ with an order of rate of convergence 3 our proof is complete. It follows from Theorem 7, that there is no loss in generality if we only consider equations with constant coefficients. The general first order difference equation of 17th degree is

$$S_{n+1} = -\frac{\sum_{j=0}^{17} a_j S_n^j}{\sum_{j=0}^{16} b_j S_n^j}.$$

If this equation defines sequences which converge to $\beta^{1/17}$ with an order of convergence 3, when we set $S_n = \beta^{1/17} + \epsilon$ and expand the right member in powers of ϵ, the constant term will be equal to $\beta^{1/17}$ and the coefficients of ϵ and ϵ^2 equal to zero, viz.

$$a_0 + (a_1 + b_0)\beta^{1/17} + (a_2 + b_1)\beta^{2/17} + \cdots + (a_{17} + b_{16})\beta = 0$$

$$a_1 + (2a_2 + b_1)\beta^{1/17} + (3a_3 + 2b_2)\beta^{2/17} + \cdots + (17a_{17} + 16b_{16})\beta^{16/17} = 0$$

$$a_2 + (3a_3 + b_2)\beta^{1/17} + (6a_4 + 3b_3)\beta^{2/17} + \cdots + (136a_{17} + 120b_{16})\beta^{15/17} = 0$$

If the a's and b's are to have rational values they must satisfy the following equations

$$a_1 + b_0 = 0, \quad a_2 + b_1 = 0, \quad \cdots, a_{16} + b_{15} = 0, \quad a_0 + (a_{17} + b_{16})\beta = 0$$

$$a_1 = 0, \quad 2a_2 + b_1 = 0, \quad 3a_3 + 2b_1 = 0, \cdots, \quad 17a_{17} + 16b_{16} = 0$$

$$a_2 = 0, \quad 3a_3 + b_2 = 0, \quad 6a_4 + 3b_3 = 0, \cdots, \quad 136a_{17} + 120b_{16} = 0,$$

which have the unique solution $a_i = 0$, $b_j = 0$ $(i = 0, 1, \cdots, 17)$ $(j = 0, 1, \cdots, 16)$. Therefore no such equation exists. Q.E.D.

Next we show that the sequences defined by a first order difference equation may converge to an algebraic number more rapidly than those defined by all other difference equation of higher order and of the same degree.

THEOREM 9. *The difference equation of second degree that defines sequences which converge most rapidly to the irrational number* $\beta^{\frac{1}{2}}$ *where* β *is rational, is the one of first order.*

PROOF. Consider the general second degree difference equation of the m^{th} order

$$\sum_{r=0}^{l}\left[\sum_{j=0}^{m-1}\sum_{i\geqq j}^{m} c_{rij}n^{-r}S_{n+j}S_{n+i} + \sum_{i=0}^{m} c_{ri}n^{-r}S_{n+i} + n^{-r}c_r\right] = 0. \tag{34}$$

When it defines sequences that approach $\beta^{\frac{1}{2}}$,

$$\sum_{j=0}^{m-1}\sum_{i\geqq j}^{m} c_{0ij}\beta + \sum_{i=0}^{m} c_{0i}\beta^{\frac{1}{2}} + c_0 = 0. \tag{35}$$

and the transformed equation is

$$\begin{aligned}\sum_{r=0}^{l} n^{-r}\left[\sum_{j=0}^{m-1}\sum_{i\geqq j}^{m} c_{rij}(\bar{S}_{n+j}\bar{S}_{n+i} + \beta^{\frac{1}{2}}\bar{S}_{n+j} + \beta^{\frac{1}{2}}\bar{S}_{n+i}) + \sum_{i=0}^{m} c_{ri}\bar{S}_{n+i}\right]& \\ + \sum_{r=1}^{l} n^{-r}\left[\sum_{j=0}^{m-1}\sum_{i\geqq j}^{m} c_{rij}\beta + \sum_{i=0}^{m} c_{ri}\beta^{\frac{1}{2}} + c_r\right] &= 0\end{aligned} \tag{36}$$

If a difference equation of the form (36) defines sequences which converge to zero with the maximum rate, then, according to Theorem 5

$$\sum_{j=0}^{m-1}\sum_{i\geqq j}^{m} c_{rij}\beta + \sum_{i=0}^{m} c_{ri}\beta^{\frac{1}{2}} + c_r = 0 \qquad (r = 1, 2, \cdots, l) \tag{35'}$$

and

$$\sum_{i\geqq j}^{m} \epsilon_{ij}c_{rji}\beta^{\frac{1}{2}} + c_{rj} = 0 \qquad (j = 0, 1, \cdots, m-1)(r = 0, 1, \cdots, l) \tag{37}$$

where

$$\epsilon_{ij} = \begin{cases}1 & i \neq j \\ 2 & i = j.\end{cases}$$

Therefore when $\beta^{\frac{1}{2}}$ is an irrational number and the coefficients of the difference equation are rational numbers, $c_{rj} = 0$, $(j = 0, 1, \cdots m - 1)$ $(r = 0, 1, \cdots, l)$ And since from (35) and (35′) $\sum_{i=0}^{m} c_{ri} = 0$ $(r = 0, 1, \cdots, l)$, $c_{rm} = 0$ $(r = 0, 1, \cdots, l)$.

Now assume our theorem is false, then there is a difference equation of the form (34) of higher than first order that defines sequences which approach $\beta^{\frac{1}{2}}$ with an order of convergence 2. Writing (36) in the form

$$\bar{S}_{n+m} = -\frac{\sum_{r=0}^{l} \frac{1}{n^r} \sum_{j=0}^{m-1} \sum_{i \geqq j}^{m-1} c_{rij}(\bar{S}_{n+j}\bar{S}_{n+i} + \beta^{\frac{1}{2}}\bar{S}_{n+j}, + \beta^{\frac{1}{2}}\bar{S}_{n+i})}{\sum_{r=0}^{l} \frac{1}{n^r} \sum_{j=0}^{m-1} c_{rmj}(\bar{S}_{n+j} + \beta^{\frac{1}{2}})},$$

we can see that if $\bar{S}_{n+m} = \theta_m \epsilon^{2^m}$ when $S_{n+i} = \theta_i \epsilon^{2^i}$ $(i = 0, 1, \cdots, m-1)$

$$c_{rij} = 0 \qquad (i, j = 0, 1, \cdots, m-2)(r = 0, \cdots, l)$$

$$c_{r,m-2,m-1} = 0 \qquad (r = 0, 1, 2, \cdots, l).$$

These results combined with the equation (37) and (35) give

$$c_{rmj} = 0 \quad (j = 0, 1, \cdots, m-2)(r = 0, 1, \cdots, l)$$

and

$$(c_{r,m-1,m} + c_{r,m-1,m-1})\beta + c_r = 0 \qquad (r = 0, 1, \cdots, l)$$

$$2c_{r,m-1,m-1} + c_{r,m-1,m} = 0 \qquad (r = 0, 1, \cdots, l).$$

Therefore the unique equation of the form (34) which defines sequences that approach $\beta^{\frac{1}{2}}$ with an order of convergence 2 is

$$2S_{n+m}S_{n+m-1} - S_{n+m-1}^2 - \beta = 0$$

but this equation is equivalent to the first order difference equation

$$2S_{n+1}S_n - S_n^2 - \beta = 0 \qquad \text{Q.E.D.}$$

In summarizing we can only say that, in general, there is no direct relation between the order of a difference equation and the order of rate of convergence of its sequences.

4. Minimum degree for a first order difference equation whose sequences approach $\beta^{1/p}$ with a given order of convergence.

THEOREM 10. *There exists a first order difference equation of q^{th} degree, $q = kp$, that define sequences which converge to $\beta^{1/p}$, where β is rational and $\beta^{j/p}$ is irrational $(j = 1, 2, \cdots, p-1)$, with an order of rate of convergence $2k$, and no first order difference equation of degree kp may define sequences which converge to $\beta^{1/p}$ with an order of convergence greater than $2k$.*

PROOF. The transformed equation of the first order difference equation of q^{th} degree,

$$S_{n+1} = \frac{-\sum_{j=0}^{q} a_j S_n^j}{\sum_{j=0}^{q-1} b_j S_n^j} \tag{38}$$

is

$$(39)\qquad \bar{S}_{n+1} = -\frac{\sum_{j=0}^{q} a_j(\bar{S}_n - \beta^{1/p})^j}{\sum_{j=0}^{q-1} b_j(\bar{S}_n - \beta^{1/p})^j} - \beta^{1/p}.$$

If the coefficients a_j, b_i $(j = 0, 1, \cdots, m)$ $(i = 0, 1, \cdots, m - 1)$ may be determined so that the coefficients of $\bar{S}_n^t$ in the numberator of the left member of (39) vanish $(t = 0, 1, \cdots, r)$, then (38) defines sequences which approach $\beta^{1/p}$ with an order of convergence $r + 1$. Setting the coefficients of $\bar{S}_n^j$ $(t = 0, 1, \cdots, r)$ equal to zero we obtain the following linear homogeneous equation in the a's and b's:

$$\sum_{j=0}^{q-1} (a_j\beta^{j/p} + b_j\beta^{(j+1)/p}) + a_q\beta^{q/p} = 0$$

$$(40)\qquad \sum_{j=0}^{q-1} j(a_j\beta^{(j-1)/p} + b_j\beta^{j/p}) + qa_q\beta^{(q-1)/p} = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_{j=r}^{q-1} \binom{j}{r} (a_j\beta^{(j-r)/p} + b_j\beta^{(j-r+1)/p}) + \binom{q}{r} \beta^{(q-p)/p} a_q = 0.$$

Since by hypothesis $\beta^{j/p}$ is irrational $(j = 1, 2, \cdots, p - 1)$, and since the a's and b's are to be rational each of the equation (40) divide into p equations. Therefore, the demand that the sequences approach $\beta^{1/p}$ with an order of convergence $2k$ imposes $2kp$ linear homogeneous relations upon the $2kp + 1$ coefficients of (38). Hence there is always a solution.

Assume that the second part of the theorem is false, then the a's and b's will satisfy the relations (40) when $r = 2k$, that is to say, they will satisfy $(2k + 1)p$ linear homogeneous equations. These $(2k + 1)p$ equations divide into p systems of equation. One with $2k + 1$ unknowns,

$$(41)\qquad \begin{cases} a_0 + \beta b_{p-1} + \beta a_p + \beta^2 b_{2p-1} + \beta^2 a_{2p} + \cdots + \beta^k b_{kp-1} + \beta^k a_q = 0 \\ \binom{p-1}{1} \beta b_{p-1} + \binom{p}{1} \beta a_p + \binom{2p+1}{1} \beta^2 b_{2p-1} + \binom{2p}{1} \beta^2 a_{2p} + \cdots \\ \qquad + \binom{kp-1}{1} \beta^k b_{kp-1} + \binom{q}{1} \beta^k a_q = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \binom{p-1}{2k} \beta b_{p-1} + \binom{p}{2k} \beta a_p + \binom{2p+1}{2k} \beta^2 b_{2p-1} + \binom{2p}{2k} \beta^2 a_{2p} + \cdots \\ \qquad + \binom{kp-1}{2k} \beta^k b_{kp-1} + \binom{q}{2k} \beta^k a_q = 0 \end{cases}$$

and $(p - 1)$ of the form

$$(42)\quad \begin{cases} b_j + a_{j+1} + \beta b_{j+p} + \beta a_{j+p+1} + \cdots + \beta^{k-1} b_{j+(k-1)p} + \beta^k a_{j+(k-1)p+1} = 0 \\ \binom{j}{1} b_j + \binom{j+1}{1} a_{j+1} + \binom{j+p}{1} \beta b_{j+p} + \binom{j+p+1}{1} \beta a_{j+p+1} \\ \qquad + \cdots + \binom{j+(k-1)p}{1} \beta^{k-1} b_{j+(k-1)p} \\ \qquad\qquad + \binom{j+(k-1)p+1}{1} \beta^k a_{j+(k-1)p+1} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \binom{j}{2k} b_j + \binom{j+1}{2k} a_{j+1} + \binom{j+p}{2k} \beta b_{j+p} + \binom{j+p+1}{2k} \beta a_{j+p+1} \\ \qquad + \cdots + \binom{j+(k-1)p}{2k} \beta^{k-1} b_{j+(k-1)p} \\ \qquad\qquad + \binom{j+(k-1)p+1}{2k} \beta^k a_{j+(k-1)p+1} = 0 \end{cases}$$

$(j = 0, 1, 2, \cdots, p - 2)$. When $j = 0$ the last equation of the system may not appear but in all cases there will be at least $2k$ equations in each of the last $p - 1$ systems. Therefore, since all of the systems of equations have at least as many equations as unknowns there are no non-trivial solutions unless a determinant of one of the systems is zero. The determinants of the coefficients are all of the form

$$\begin{vmatrix} \binom{j}{0} & \binom{j+1}{0} & \binom{j+p}{0} & \binom{j+p+1}{0} & \cdots & \binom{j+mp}{0} & \binom{j+mp+1}{0} \\ \binom{j}{1} & \binom{j+1}{1} & \binom{j+p}{1} & \binom{j+p+1}{1} & \cdots & \binom{j+mp}{1} & \binom{j+mp+1}{1} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \binom{j}{r} & \binom{j+1}{r} & \binom{j+p}{r} & \binom{j+p+1}{r} & \cdots & \binom{j+mp}{r} & \binom{j+mp+1}{r} \end{vmatrix}.$$

Such a determinant is equal to

$$\frac{1}{2!3!\cdots r!} \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ j & j+1 & j+p & \cdots & j+mp+1 \\ j^2 & (j+1)^2 & (j+p)^2 & \cdots & (j+mp+1)^2 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ j^r & (j+1)^r & (j+p)^r & \cdots & (j+mp+1)^r \end{vmatrix},$$

which is different from zero, for it is a Vandermonde determinant. Therefore, we have a contradiction. Q.E.D.

It is not difficult to see that the equation of q^{th} degree, $q = kp$, which defines sequences that converge to $\beta^{1/p}$ with an order of convergence $2k$ is unique. The solutions of the first $2k$ equations of the last $p - 1$ systems (42) are all zero and the solutions of the first $2k$ equations are of the form $c_i k$. Hence, when the factor k is removed from the numerator and the denominator, we obtain the equation.

THEOREM 11. *There is a first order difference equation of degree $kp + 1$ that defines sequences which converge to $\beta^{1/p}$, where β is rational and $\beta^{j/p}$ is irrational $(j = 1, 2, \cdots, p - 1)$, more rapidly than those defined by any first order difference equation of degree q, where $(k + 1)p > q > kp + 1$.*

PROOF. Consider the difference equation (38) for $q = (k + 1)p - 1$. First, assume that the sequences defined by this equation approach $\beta^{1/p}$ with an order of convergence $2k + 2$. Then the coefficients satisfy $2k + 2$ relations (40). These $2k + 2$ relations give $(2k + 2)p$ homogeneous linear equations in the a's and b's. And the linear equations divide into p systems of $2k + 2$ equations each. One of the systems contains $2k + 1$ unknowns and the other $p - 1$ systems each contain $2k + 2$ unknowns. None of these systems have non-trivial solutions for the determinants of the coefficients are composed of columns of binomial coefficients which we proved, in the proof of Theorem 10 to be non-vanishing. Hence our assumption is false.

Second, assume that the sequences have an order of rate of convergence $2k + 1$, then the a's and the b's must satisfy $2k + 1$ relations (40). These $2k + 1$ relations give p systems of homogeneous linear equations (41) and (42). One system contains $2k + 1$ unknowns and the other $p - 1$ systems contain $2k + 2$ unknowns each. The solutions of the first system must all be zero, while each of the other $p - 1$ systems may be solved for $2k + 1$ of their unknowns in terms of one of the others, for again, all determinants of the coefficients are different from zero.

We construct a matrix of $2(k + 1)$ columns and $(2k + 1)(p - 1)$ rows, from the coefficients of the $p - 1$ systems of equations, where the coefficients of $a_i\beta^j$, $a_{i+1}\beta^j, \cdots, a_{i+p-1}\beta^j$ lie in the same column $(i = 1, p + 1, \cdots, kp + 1)$ and the coefficients $b_l\beta^j, b_{l+1}\beta^j, \cdots, b_{l+p-1}\beta^j$ lie in the same column $(l = 0, p, 2p, \cdots, kp)$. That is, the matrix

$$\begin{Vmatrix}
1 & 1 & 1 & 1 & \cdots & 1 & 1 \\
0 & \binom{1}{1} & \binom{p}{1} & \binom{p+1}{1} & \cdots & \binom{kp}{1} & \binom{kp+1}{1} \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\
0 & 0 & \binom{p}{2k} & \binom{p+1}{2k} & \cdots & \binom{kp}{2k} & \binom{kp+1}{2k} \\
1 & 1 & 1 & 1 & \cdots & 1 & 1 \\
\binom{1}{1} & \binom{2}{1} & \binom{p+1}{1} & \binom{p+2}{1} & \cdots & \binom{kp+1}{1} & \binom{kp+2}{1} \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\
0 & \binom{2}{2k} & \binom{p+1}{2k} & \binom{p+2}{2k} & \cdots & \binom{kp+1}{2k} & \binom{kp+2}{2k} \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\
1 & 1 & 1 & 1 & \cdots & 1 & 1 \\
\binom{p-2}{1} & \binom{p-1}{1} & \binom{2p-2}{1} & \binom{2p-1}{1} & \cdots & \binom{(k+1)p-2}{1} & \binom{(k+1)p-1}{1} \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\
\binom{p-2}{2k} & \binom{p-1}{2k} & \binom{2p-2}{2k} & \binom{2p-1}{2k} & \cdots & \binom{(k+1)p-2}{2k} & \binom{(k+1)p-1}{2k}
\end{Vmatrix}$$

Since the binomial coefficients obey the law

$$\binom{r}{n-1} + \binom{r}{n} = \binom{r+1}{n}$$

any one of the rows of this matrix is linearly dependent upon the first $2k + 1$ rows. Hence the rank of the matrix is $2k + 1$.

This tells us that if we solve the first system of equations for $2k + 1$ of the unknowns in terms of a_i that the second set of equations has the same solutions where the subscripts are increased by one, and so on. If we solve $p - 1$ systems of equations in terms of $a_{kp+1}, a_{kp+2}, \cdots, a_{(k+1)p-1}$ respectively, then the given equation (38) takes the form

$$S_{n+1} = \frac{\phi_1(S_n) \sum_{j=0}^{p-1} a_{kp+j} S_n^{j-1}}{\phi_2(S_n) \sum_{j=1}^{p-1} a_{kp+j} S_n^{j-1}}$$

or

$$S_{n+1} = \frac{\phi_1(S_n)}{\phi_2(S_n)}, \tag{43}$$

where $\phi_1(S_n)$ and $\phi_2(S_n)$ are of degrees $kp + 1$ and kp respectively. Since the equation (43) is unique, our proof is complete. Q.E.D.

The two previous theorems have dealt only with difference equations with constant coefficients. In the following theorem we assume that the coefficients are polynomials in n.

THEOREM 12. *There exists some first order difference equations with constant coefficients that define rational sequences which converge to $\beta^{1/p}$, where $\beta^{j/p}$ is irrational $(j = 1, 2, \cdots, p - 1)$ more rapidly than the sequences defined by any first order difference equation whose coefficients are polynomials in n.*

PROOF. If the difference equation

$$S_{n+1} = -\frac{\sum_{i=0}^{l} n^i \sum_{j=0}^{q} a_{ij} S_n^j}{\sum_{i=0}^{l} n^i \sum_{j=0}^{q-1} b_{ij} S_n^j} \tag{44}$$

defines sequences which converge to $\beta^{1/p}$ with an order of convergence λ, then

$$\begin{aligned}
&\sum_{j=0}^{q-1} (a_{ij}\beta^{j/p} + b_{ij}\beta^{(j+1)/p}) + a_{iq}\beta^{q/p} = 0 \qquad (i = 0, 1, \cdots, l)\\
&\sum_{j=0}^{q-1} j(a_{ij}\beta^{(j-1)/p} + b_{ij}\beta^{j/p}) + q a_{iq}\beta^{(q-1)/p} = 0 \qquad (i = 0, 1, \cdots, l)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
&\sum_{j=0}^{q-1} \binom{j}{r-1} [a_{ij}\beta^{(j-r+1)/p} + b_{ij}\beta^{(j-r)/p}]\\
&\qquad\qquad + \binom{q}{r-1} a_{iq}\beta^{(q-r+1)/p} = 0, \quad (i = 0, 1, \cdots, l).
\end{aligned} \tag{45}$$

When $l = 0$ we know from the previous theorems that there is a maximum order λ for the rate of convergence for the sequences. That is, the number of relations for the determination of a_{0j} and b_{0j} cannot exceed λ. The relations (45) are similar for all values of i. Hence when $l > 0$, the number of such relations cannot exceed λ, without all a_{ij} and b_{ij} vanishing.

When $q = kp$ we know, after Theorem 10 that $\lambda = 2k$. Each of these $2k$ relations (45) gives, for each i, q linear homogeneous equations on the coefficients. Since the relations are the same for all i, when the last system of equations are solved in terms of a_{ir} $(i = 0, 1, \cdots, l)$ (45) reduces to

$$S_{n+1} = \frac{\phi_1(S_n) \sum_{i=0}^{l} n^i a_{ir}}{\phi_2(S_n) \sum_{i=0}^{l} n^i a_{ir}}$$

or

$$S_{n+1} = \frac{\phi_1(S_n)}{\phi_2(S_n)},$$

where ϕ_1 and ϕ_2 are polynomials is S_n with constant coefficients.

When $q = (k + 1)p - 1$, then as in the proof of Theorem 11 for $l = 0$

$$S_{n+1} = \frac{\phi_3(S_n) \sum_{j=0}^{p-1} a_{0kp+j} S_n^{j-1}}{\phi_4(S_n) \sum_{j=0}^{p-1} a_{0kp+j} S_n^{j-1}}$$

and since the relations (45) are similar for all i

$$S_{n+1} = \frac{\phi_3(S_n) \sum_{i=0}^{l} \sum_{j=0}^{p-1} a_{ikp+j} S_n^{j-1}}{\phi_4(S_n) \sum_{i=0}^{l} \sum_{j=0}^{p-1} a_{ikp+j} S_n^{j-1}},$$

or

$$S_{n+1} = \frac{\phi_3(S_n)}{\phi_4(S_n)},$$

where ϕ_3 and ϕ_4 are polynomials with constant coefficients. Q.E.D.

EXAMPLE: When $k = 1$, the first order difference equation of p^{th} degree whose sequences converge most rapidly to $\beta^{1/p}$, where β is rational and $\beta^{j/p}$ is irrational $(j = 1, 2, \cdots, p - 1)$, is

$$pS_{n+1}S_n^{p-1} - (p - 1)S_n^p - \beta = 0 \tag{46}$$

and the one of $(p + 1)^{\text{st}}$ degree is

$$(p + 1)S_{n+1}S_n^p - (p - 1)S_n^{p+1} + (p - 1)\beta S_{n+1} - (p + 1)S_n\beta = 0. \tag{47}$$

The sequences defined by (46) have an order of convergence 2 and those defined by (47) have an order of convergence 3.

It is interesting to note that equation (46) is *Newton's algorithm for the approximation of* $\beta^{1/p}$.

5. Approximation of $\beta^{i/p}$. In practical work it is often valuable to know an upper bound for the error of a given term of a sequence, i.e., an upper bound for the difference between a term and the limit of the sequence, or in other words, if the terms of the sequence are written in decimal form it is important to know a means of asserting that a given term of the sequence is correct to at least R decimal places.

THEOREM 13. *If a sequence approaches the real number* $\beta^{1/p}$, *such that when* $|S_n - \beta^{1/p}| = \epsilon$ *then* $|S_{n+1} - \beta^{1/p}| \leqq C\epsilon^k$, *where* C *is a constant independent of* ϵ, (i.e. the sequence has an order of convergence k) *and if,* ϵ *is so small that* $2C\epsilon^k < \epsilon$, *and if the first* i *decimal places of the two consecutive terms of the sequence are the same,* $(i > 0)$; *then, the second of the terms cannot differ from* $\beta^{1/p}$ *by as much as* $2^k(10)^{-ki}C$. *And when* $C < (10)^j$ *the second of the terms is correct to* $ki - j - k/3$ *decimal places.*

PROOF. Let ϵ_n denote the error of S_n. Since $C\epsilon_n^{k-1} < \frac{1}{2}$,

$$\epsilon_n < 2\epsilon_n |1 - C\epsilon_n^{k-1}|$$
$$= 2|\epsilon_n - C\epsilon_n^k| \leqq 2||S_n - \beta^{1/p}| - |S_{n+1} - \beta^{1/p}|| \leqq 2|S_n - S_{n+1}|.$$

Now

$$|S_n - S_{n+1}| < (10)^{-i},$$

hence

$$\epsilon_{n+1} \leqq C\epsilon_n^k < C2^k(10)^{-ki}.$$

And when

$$C < 10^j$$
$$\epsilon_{n+1} < 2^k(10)^{j-ki} < 10^{k(-i+\frac{1}{3})+j}.$$

Therefore, S_{n+1} is correct to at least $ki - j - k/3$ decimal places.

EXAMPLE: Consider the sequences defined by (27). For this equation $k = 2$, $|C| < 1$ therefore $j = 0$. For the sequence with initial condition $S_0 = 1$, S_3 and S_4 differ only in the sixth decimal place, hence S_4 is correct to 9 decimal places. By the same method of calculation S_5 is correct to 21 decimal places.

Although it is possible to set up a difference equation whose sequences converge as rapidly as one desired, there seems to be less calculation in the approximation of an irrational number $\beta^{1/p}$ if the sequences are defined by the simple difference equation (46) *Newton's algorithm*, which may be written in the form

$$S_{n+1} = \frac{p-1}{p} S_n + \frac{\beta}{pS_n^{p-1}}. \tag{48}$$

We know that the order of rate of convergence of its sequences is 2. Now let us determine its region of strong convergence.

THEOREM 14. *If $\beta^{1/p} > 0$, the sequences defined by* (48) *converge to $\beta^{1/p}$ for all positive initial values S_0. Moreover, they approach $\beta^{1/p}$ monotonically when the initial values are $> \beta^{1/p}$.*

PROOF. The transformation equation is

$$p(\bar{S}_{n+1} + \beta^{1/p})(\bar{S}_n + \beta^{1/p})^{p-1} - (p - 1(\bar{S}_n + \beta^{1/p})^p - \beta = 0$$

or

$$\bar{S}_{n+1} = \bar{S}_n\left[1 - \frac{1}{p} - \frac{\beta^{1/p}(\bar{S}_n + \beta^{1/p})^{p-1} - \beta}{p\bar{S}_n(\bar{S}_n + \beta^{1/p})^{p-1}}\right].$$

For positive values of $\bar{S}_n$ the expression in the brackets is > 0 and $< 1 - 1/p$. Therefore, the sequences S_n decrease monotonically to zero for initial values $\bar{S}_0 > 0$. This is equivalent to the statement that all sequences $\{S_n\}$ of (48) determined by initial conditions $> \beta^{1/p}$ converge monotonically to $\beta^{1/p}$. Hence the second part of the theorem is true.

If $0 < S_n < \beta^{1/p}$, then $S_{n+1} - S_n = (1/p)\{-S_n + \beta S_n^{1-p}\} > 0$, so that $S_{n+1} > S_n$. Hence, either the sequence $\{S_n\}$ is monotone increasing tending to a limit $\leqq \beta^{1/p}$, or some term of the sequence is greater than $\beta^{1/p}$ and from this term on the sequence decreases monotonically to $\beta^{1/p}$. In the first case the sequence cannot approach a limit K, $0 < K < \beta^{1/p}$ for the only finite limits for the sequence are $\beta^{1/p}$ and $-\beta^{1/p}$ since the limit of the difference equation is $S_0^p = \beta$. Hence our proof is complete. Q.E.D.

Since irrational numbers of the form $\beta^{1/p}$ are usually accurately computed by means of a binomial series, it is interesting to compare the rate of convergence of the sequence of partial sums of a binomial series with that of a sequence defined by a difference equation.

THEOREM 15. *A sequence of partial sums of a binomial series cannot converge to $\beta^{1/p}$ with a order of convergence* 2.

PROOF. CASE I. Positive term series, $(1 - 1/a)^l = 1 + b_1 + b_2 + \cdots + b_n + \cdots$. Assume that the theorem is false, then the sequence of partial sums has an order of rate of convergence 2; i.e. if

$$\beta^{1/p} - S_n = \epsilon_n \quad \text{and} \quad \beta^{1/p} - S_{n+1} = \epsilon_{n+1} \quad \text{then} \quad \epsilon_{n+1} < C\epsilon_n^2,$$

where C is a constant independent of ϵ_n. For a rapidly convergent series $b_{n+1}/b_n < r < 1$ for $n > n_0$. The error ϵ_n of the partial sum S_n is less than $b_n\left(\frac{r}{1-r}\right)$ and is greater than b_{n+1}. Therefore

$$\epsilon_n < b_n\left(\frac{r}{1-r}\right) = \binom{l}{n}\frac{1}{(-a)^n}\cdot\left(\frac{r}{1-r}\right),$$

$$\binom{l}{n+1}\left(\frac{1}{-a}\right)^{n+1} = b_{n+1} < \epsilon_{n+1}.$$

Therefore, if the theorem is false,

$$\binom{l}{n+1}\left(\frac{1}{-a}\right)^{n+1} < C\binom{l}{n}^2 \frac{1}{a^{2n}}\left(\frac{r}{1-r}\right)^2$$

or

$$\left|\frac{(l-n)}{(n+1)}\frac{1}{a}\right| < C\left|\frac{l(l-1)\cdots(l-n+1)}{n!}\frac{1}{a^n}\right|\left(\frac{r}{1-r}\right)^2.$$

This last inequality cannot hold for all n, if $a > 1$. But, since the series is convergent, $a > 1$. Hence we have a contradiction.

Case II. Alternating series. $(1 + 1/a)^l = 1 - b_1 + b_2 - b_3 + \cdots$. Again assume the theorem is false. The error ϵ_n is less than b_{n+1} and is greater than $b_{n+1} - b_{n+2}$. So

$$\epsilon_n < b_{n+1} = \left|\binom{l}{n+1}\frac{1}{a^{n+1}}\right|$$

$$\left|\binom{l}{n+2}\frac{1}{a^{n+2}} - \binom{l}{n+3}\frac{1}{a^{n+3}}\right| = b_{n+2} - b_{n+3} < \epsilon_{n+1}.$$

Therefore if the theroem is false

$$\left|\binom{l}{n+2}\frac{1}{a^{n+2}} - \binom{l}{n+3}\frac{1}{a^{n+3}}\right| < C\binom{l}{n+1}^2\left(\frac{1}{a^{n+1}}\right)^2$$

$$\left|\frac{l-n-1}{n+2}\cdot\frac{a - \dfrac{l-n-2}{n+3}}{a^2}\right| < C\left|\binom{l}{n+1}\frac{1}{a^{n+1}}\right|.$$

This last inequality cannot hold for all n, for the left member approaches a constant greater than zero, ($a \neq 1$), and the right member approaches zero since it is a term of the series. Therefore we have a contradiction. Q.E.D.

Corollary: *There are sequences defined by the difference equation* (48) *which converge more rapidly to* $\beta^{1/p}$ *than the sequence of partial sums of any binomial series.*

University of Maryland
College Park, Maryland

REFERENCES

(1) S. Lattès, "Sur les suites récurrentes non linéaires et sur les fonctions génératrices de ces suites," *Ann. Fac. Sc. Toulouse* (3), 3 (1911), pp. 73-124.

(2) J. Horn, "Zur Theorie der nicht linearen Differential- und Differenzengleichungen," *J. für Math.* 141, (1912), pp. 182-216.

(3) O. Perron, "Über Stabilität und asymptotisches Verhalten der Lösungen eines Systems endlicher Differenzengleichungen," *J. für Math.* 161, (1929), pp. 41-64.

(4) E. Schröder, "Über unendlich viele Algorithmen zur Auflösung der Gleichungen," *Math. Ann.* 2, (1870), pp. 317-365.

Annals of Mathematics
Vol. 42, No. 1, January, 1941

SOME GENERAL APPROXIMATION THEOREMS

By Lawrence M. Graves

(Received February 3, 1940)

1. Introduction. In this paper comparatively simple proofs are given for some general theorems on the approximation of a given function by "smoother" functions. Well known theorems on approximation by polynomials, by trigonometric sums, and by integral means are special cases. The proofs, given in Section 3, are so formulated that there is little more complication in the case of functions of m variables than in the case of functions of one variable. Theorems on the term-by-term integration and differentiation of the approximating sequence are included. In Section 4 certain cases of iteration of the approximating transformations are considered. In Section 5 it is shown that when the given function is continuous and vanishes on a closed point set E, certain types of approximating functions may be required to vanish on a neighborhood of E while possessing as many continuous derivatives as may be derived.

Let A denote the interval $0 \leqq x_i \leqq a_i$, $(i = 1, \cdots, m)$, of m-dimensional space, and let B denote the interval $-a_i \leqq x_i \leqq a_i$. Let $\{Q_n(x)\}$ be an infinite sequence of bounded measurable functions defined on B and let (k_n) be an infinite sequence of positive numbers. Then if $f(x)$ is a function defined and integrable on A, we may set

$$\begin{aligned} P_n(x;f) &= k_n \int_A f(z)Q_n(z-x)\,dz \\ &= k_n \int_0^{a_1} \cdots \int_0^{a_m} f(z)Q_n(z-x)\,dz_m \cdot \cdot\, dz_1. \end{aligned} \tag{1:1}$$

It is observed at once that if the function $Q_n(x)$ is of class $C^{(r)}$ on B then P_n is of class $C^{(r)}$ on A, and if $Q_n(x)$ is a polynomial so is P_n. For the most part we shall wish to assume that the numbers k_n are related to the functions Q_n by the formula

$$1/k_n = \int_B Q_n(x)\,dx = \int_{-a_1}^{a_1} \cdots \int_{-a_m}^{a_m} Q_n\,dx_m \cdots dx_1. \tag{A}$$

Interesting examples of sequences $Q_n(x)$ are:

(i) $Q_n(x) = [1 - \sum_i x_i^2]^n$, with $\sum_i a_i^2 \leqq 1$;

(ii) $Q_n(x) = \prod_i (1 - x_i^2)^n$, with $a_i \leqq 1$;

(iii) $Q_n(x) = \prod_i \cos^n x_i$, with $a_i \leqq \pi/2$;

(iv) $\quad Q_n(x) = 1 \quad \text{for} \quad |x_i| \leqq 1/n, \qquad Q_n(x) = 0 \quad \text{elsewhere};$

(v) $\quad Q_n(x) = \prod_i (1/n^2 - x_i^2)^n \quad \text{for} \quad |x_i| \leqq 1/n, \qquad Q_n(x) = 0 \quad \text{elsewhere}.$

In examples (i) and (ii) the polynomials $P_n(x)$ related to a function f by formula (1:1) are frequently called Stieltjes polynomials. In example (iii) we obtain trigonometric polynomials $P_n(x)$. In example (iv) the functions $P_n(x)$ are integral means of $f(x)$ over hyper-cubes.

We shall wish to consider the following properties which sequences (Q_n) and (k_n) may possess. We let S_ϵ denote the interior of the sphere with center at the origin and radius ϵ.

(B) *Each $Q_n(x)$ is bounded and measurable and non-negative on the domain B, and positive on a subset of B of positive measure.*

(C) *For every $\epsilon > 0$, $\lim_n k_n Q_n(x) = 0$ uniformly on $B - S_\epsilon$.*

(D) *Each Q_n is of class $C^{(r)}$ on B.*

(E) *If R_n denotes a partial derivative of Q_n, of total order less than r, then for every $\epsilon > 0$, $\lim_n k_n R_n(x) = 0$ uniformly on $B - S_\epsilon$.*

When the numbers k_n are given by formula (A), examples (i), (ii) and (iii) possess all these properties, and so does example (*v*) for $n > n_r$ sufficiently great. Example (iv) has properties (*B*) and (*C*). To verify property (*E*) for example (i), we notice that if R_n is a derivative of Q_n of order μ, then R_n is $O[n^\mu(1 - \epsilon^2)^n]$ uniformly on $B - S_\epsilon$. Also, with $\delta = n^{-1/2}$,

$$1/k_n \geqq \int_{S_\delta} [1 - \sum x_i^2]^n \, dx \geqq \int_{S_\delta} [1 - 1/n]^n \, dx = bn^{-m/2}[1 - 1/n]^n,$$

where b is a properly chosen constant, so that k_n is $O(n^{m/2})$.

In case the function f is originally defined only on a subset A_0 of A, its domain of definition may be extended to be the whole of A in a variety of ways. If A_0 is the interval $b_i \leqq x_i \leqq c_i$, we may set

$$\begin{gathered} f(x_1, x_2, \cdots, x_m) = f(b_1, x_2, \cdots, x_m), \\ (0 \leqq x_1 < b_1, \qquad b_i \leqq x_i \leqq c_i, \qquad i = 2, \cdots, m), \end{gathered} \tag{1:2}$$

and so on until the extension is completed. With this extension the properties of continuity, of m-tuple absolute continuity, and of satisfying a Lipschitz condition are all preserved. In place of (1:2) we might use

$$f(x_1, x_2, \cdots, x_m) = f(2b_1 - x_1, x_2, \cdots, x_m). \tag{1:3}$$

This extension preserves also the property of absolute continuity in the sense on Tonelli. If f is integrable on a measurable subset A_0 of A, we may set $f = 0$ on the complement of A_0. For an extension preserving uniform continuity or a Lipschitz condition when A_0 is an arbitrary subset of A, the reader is referred to McShane[1] [10] or Whitney [11, p. 63, footnote]. When f is of

[1] Numbers in brackets refer to the bibliography at the end.

class $C^{(r)}$ (in a suitable sense) on a closed set A_0, an extension preserving this property has been given by Whitney [11]. A simpler method of extension applicable in special cases has been devised by Hestenes, but is not yet published.

We shall find it convenient throughout the proofs to set $f = 0$ outside the interval A. Then by making the change of variables $z_i = x_i + v_i$ we find that for x in A the formula (1:1) may be written

$$P_n(x;f) = k_n \int_B f(x+v)Q_n(v)\,dv. \tag{1:4}$$

2. Preliminary lemmas. The class of all measurable functions f such that $|f|^p$ is integrable on A is denoted as usual by L^p, for $1 \leqq p < \infty$. For this class set

$$||f||_p = \left[\int_A |f|^p\,dx\right]^{1/p}. \tag{2:1}$$

We make the convention that L^∞ denotes the class of essentially bounded and measurable functions, and for this class denote by $||f||_\infty$ the essential upper bound of $|f(x)|$ on A. Then if $p + q = pq$ ($q = \infty$ when $p = 1$) we have the Hölder inequality

$$\left|\int_A f(x)g(x)\,dx\right| \leqq ||f||_p \cdot ||g||_q \tag{2:2}$$

holding when f is in L^p and g is in L^q, and the triangle inequality[2]

$$||f + g||_p \leqq ||f||_p + ||g||_p \tag{2:3}$$

holding when f and g are both in L^p.

We shall have need for two lemmas, which are readily proved.

LEMMA 1. *Let $f(x)$ be in L^p, $(1 \leqq p < \infty)$, and set $f_h(x) = f(x + h)$. Then $||f_h - f||_p$ approaches zero with h.*

To prove this note that by the triangle inequality (2:3), $||f_h - f||_p \leqq ||f_{1h} - f_1||_p + ||f_{2h} - f_2||_p$ when $f = f_1 + f_2$, and $||f_h - f||_p \leqq 2\,||f||_p$. From these inequalities it follows that the class E of functions f for which the conclusion of the lemma holds is linear and closed in L^p. The class E obviously contains the characteristic functions of sub-intervals of A. The class of all linear combinations of such characteristic functions is dense[3] on L^p, and hence $E = L^p$.

LEMMA 2. *If $f(x)$ is in L^p, $(1 \leqq p \leqq \infty)$, on A, then $f(x + v)$ as a function of $2m$ variables, x in A and v in B, is also in L^p.*[4]

To prove this we note first that if a set E is closed in A, then the set E^* of points (x, v) of the product space AB for which $x + v$ is in E is likewise closed

[2] See, e.g., Hobson [8], vol. I, p. 588.
[3] See, e.g., Hobson [8], vol. II, p. 251.
[4] Compare Cinquini [6], pp. 59–62.

in AB. Consequently if E is open in A, E^* is open in AB. If E is an interval with measure μE, then E^* has measure

$$\mu E^* = \mu E \cdot \prod a_i . \tag{2:4}$$

Since every open set may be represented as a sum of non-overlapping closed intervals, it follows that (2:4) holds for sets E open in A, and hence for sets E closed in A. By applying the criterion that a set E is measurable if and only if for every $\epsilon > 0$ there exists a closed set E_1 and an open set E_2 such that $E_1 \subset E \subset E_2$ and $\mu(E_2 - E_1) < \epsilon$, we see that when E is measurable E^* is also measurable, and (2:4) still holds. Hence $f(x + v)$ is measurable on AB. Let $_N(x) = |f(x)|$ where $|f(x)| \leqq N$, $f_N(x) = N$ where $|f(x)| > N$. Then

$$\int_A \int_B [f_N(x + v)]^p \, dv \, dx = \int_A \int_A [f_N(v)]^p \, dv \, dx$$

$$= \int_A [f_N(v)]^p \, dv \cdot \prod_i a_i \leqq \int_A |f(v)|^p \, dv \cdot \prod a_i .$$

This with Fubini's theorem[5] implies the desired conclusion.

3. The general approximation theorems. For the sake of completeness, we include the following theorem, whose proof is an immediate extension of that given by Landau [1].

THEOREM 1. *Suppose the sequences (Q_n) and (k_n) have properties (A), (B), and (C). Let A_0 be a closed set interior to A, and suppose that $f(x)$ is bounded and measurable on A, and continuous at the points of A_0. Then $P_n(x)$ converges uniformly to $f(x)$* on A_0.

Let $\alpha > 0$ be the minimum distance from A_0 to the boundary of A. For an arbitrary $\delta > 0$ there exists a positive $\epsilon < \alpha$ such that

$$|f(x + v) - f(x)| < \delta \tag{3:1}$$

whenever x is in A_0 and $\sum_i v_i^2 < \epsilon^2$. Now from formula (1:4) and property (A) we have

$$|P_n(x) - f(x)| = k_n \left| \int_B [f(x + v) - f(x)] Q_n(v) \, dv \right|$$

$$\leqq k_n \int_{S_\epsilon} |f(x + v) - f(x)| Q_n(v) \, dv \tag{3:2}$$

$$+ k_n \int_{B - S_\epsilon} [|f(x + v)| + |f(x)|] Q_n(v) \, dv.$$

The first term on the right is less than δ for x in A_0 by (3:1) and properties (A) and (B), and the second term approaches zero with $1/n$ uniformly in x, by property (C).

[5] Saks [9], p. 77, Hobson [8], vol. 1, p. 577.

THEOREM 2. *Suppose the sequences* (Q_n) *and* (k_n) *have properties* (A), (B), *and* (C). *Let* $f(x)$ *be in* L^p, $(1 \leqq p < \infty)$. *Then*

(a) $\lim_n \int_A |P_n(x) - f(x)|^p dx = 0$;

(b) *a subsequence* $P_{n_i}(x)$ *converges to* $f(x)$ *almost everywhere on* A;

(c) *the integrals* $\int_E |P_n(x)|^p dx$ *are equi-absolutely continuous*;

(d) *if* M *is the essential upper bound of* $|f(x)|$, *then* $|P_n(x)| \leqq M$ *on* A;

(e) *if* $g(x)$ *is in* $L^q(p + q = pq, 1 \leqq p \leqq \infty)$,

$$\lim_n \int_A |g(x)||P_n(x) - f(x)| dx = 0.$$

By the use of the first equality in (3:2) and the Hölder inequality we have

$$\begin{aligned}
|P_n(x) - f(x)|^p &= k_n^p \left| \int_B [f(x+v) - f(x)] Q_n(v)\, dv \right|^p \\
&\leqq k_n^p \int_B |f(x+v) - f(x)|^p Q_n(v)\, dv \left[\int_B Q_n(v)\, dv \right]^{p-1} \\
&= k_n \int_B |f(x+v) - f(x)|^p Q_n(v)\, dv.
\end{aligned}$$

From this by Lemma 2 and Fubini's theorem on interchange of order of integration we obtain

$$\begin{aligned}
\int_A |P_n(x) - f(x)|^p dx &\leqq k_n \int_B \int_A |f(x+v) - f(x)|^p Q_n(v)\, dx\, dv \\
&= k_n \int_{S_\epsilon} \int_A |f(x+v) - f(x)|^p dx\, Q_n(v)\, dv \\
&\quad + k_n \int_{B-S_\epsilon} \int_A |f(x+v) - f(x)|^p dx\, Q_n(v)\, dv.
\end{aligned}$$

The first term on the right converges to zero with ϵ, uniformly with respect to n, by Lemma 1 and properties (A) and (B). The second term converges to zero with $1/n$ for each ϵ, by property (C). This proves part (a) of the conclusion. Part (b) is a consequence of (a).[6] Part (c) follows by simple arguments from the triangle inequality

$$\left[\int_E |P_n(x)|^p dx \right]^{1/p} \leqq \left[\int_E |P_n(x) - f(x)|^p dx \right]^{1/p} + \left[\int_E |f(x)|^p dx \right]^{1/p},$$

part (a), and the absolute continuity of $\int_E |P_n(x)|^p dx$ and $\int_E |f(x)|^p dx$.

---

[6] See Hobson [8], vol. II, pp. 239–45.

Part (d) is an immediate consequence of (1:4). In case p is finite, part (e) follows at once from the Hölder inequality and part (a). In case f is bounded and g merely integrable, we may make an indirect proof, supposing that

$$\limsup \int_A |g(x)|\,|P_n(x) - f(x)|\,dx = \beta > 0.$$

Then for a properly selected subsequence (P_{n_j}) we have

$$\lim \int_A |g(x)|\,|P_{n_j}(x) - f(x)|\,dx = \beta,$$

$$\lim P_{n_j}(x) = f(x) \text{ almost everywhere,}$$

$$|g(x)|\,|P_{n_j}(x) - f(x)| \leqq 2M\,|g(x)| \text{ almost everywhere.}$$

This leads to a contradiction with the theorem on term-by-term integration for Lebesgue integrals.

LEMMA 3. *Suppose that the sequences* (Q_n) *and* (k_n) *have properties* (C) *and* (D) *with* $r \geqq 1$. *Let* $f(x)$ *be absolutely continuous in* x_1 *for almost all* $(x_2, \cdots, x_m)$, *and let* f *and the partial derivative* f_{x_1} *be integrable on* A. *Suppose also that* $f(0, x_2, \cdots, x_m)$ *is integrable. Let* A_0 *be a closed set interior to* A. *Then* $\lim_n |P_n(x; f_{x_1}) - P_{nx_1}(x; f)| = 0$ *uniformly in* A_0.

Since the functions Q_n are of class C′, differentiation under the integral sign is permissible in formula (1:1), and we have

$$P_{nx_1}(x; f) = -k_n \int_A f(z) Q_{nz_1}(z - x)\,dz.$$

Hence by Fubini's theorem and the absolute continuity of f in x_1 we have

$$\begin{aligned} P_n(x; f_{x_1}) - P_{nx_1}(x; f) &= k_n \int_A [f_{x_1}(z) Q_n(z - x) + f(z) Q_{nz_1}(z - x)]\,dz \\ &= k_n \int_0^{a_2} \cdots \int_0^{a_m} [f(z) Q_n(z - x)]_{z_1=0}^{z_1=a_1}\,dz_m \cdots dz_2. \end{aligned} \tag{3:3}$$

Also

$$\int_A f_{x_1}(x)\,dx = \int_0^{a_2} \cdots \int_0^{a_m} [f(x)]_{x_1=0}^{x_1=a_1}\,dx_m \cdots dx_2,$$

so that the integrability of $f(0, x_2, \cdots, x_m)$ implies that of $f(a_1, x_2, \cdots, x_m)$. Thus (3:3) with property (C) justifies the desired conclusion.

Let us recall that a function $f(x_1 \cdots x_m)$ of several variables is absolutely continuous on A in the sense of Tonelli in case it is continuous in $(x_1, \cdots, x_m)$, absolutely continuous in each x_i for almost all values of the remaining variables, and each partial derivative f_{x_i} is integrable on A.

THEOREM 3. *Suppose the sequences* (Q_n) *and* (k_n) *have properties* (A), (B), (C), *and* (D) *for* $r \geqq 1$. *Let* $f(x)$ *be absolutely continuous in the sense of Tonelli, and*

let the partial derivative f_{x_i} *be in* L^p, $(1 \leqq p < \infty)$. *Let* A_0 *be a closed set interior to* A. *Then*

(a) $\lim_n \int_{A_0} | P_{nx_i}(x) - f_{x_i}(x) |^p \, dx = 0$;

(b) *a subsequence of* (P_{nx_i}) *converges to* f_{x_i} *almost everywhere on* A;

(c) *for* $E \subset A_0$, *the integrals* $\int_E | P_{nx_i}(x) |^p \, dx$ *are equi-absolutely continuous*;

(d) *if* f *satisfies a Lipschitz condition on* A, *the partial derivatives* P_{nx_i} *are uniformly bounded on* A_0;

(e) *if* f_{x_i} *is continuous at the points of* A_0, *the sequence* (P_{nx_i}) *converges to* f_{x_i} *uniformly on* A_0;

(f) *if* $g(x)$ *is in* $L^q (p + q = pq, 1 \leqq p \leqq \infty)$,

$$\lim_n \int_{A_0} | g(x) | \, | P_{nx_i}(x) - f_{x_i}(x) | \, dx = 0.$$

To prove part (a) we use the triangle inequality

$$\left[\int_{A_0} | P_{nx_i}(x;f) - f_{x_i}(x) |^p \, dx\right]^{1/p} \leqq \left[\int_{A_0} | P_n(x;f_{x_i}) - f_{x_i}(x) |^p \, dx\right]^{1/p} + \left[\int_{A_0} | P_{nx_i}(x;f) - P_n(x;f_{x_i}) |^p\right]^{1/p}.$$

The first term on the right approaches zero by Theorem 2 with f replaced by f_{x_i}, and the second term approaches zero by Lemma 3. To prove (b) we may consider a sequence of closed intervals A_0 interior to A but with boundaries approaching the boundary of A. The existence of the desired subsequence converging almost everywhere on A is then secured by the well-known diagonal method. Parts (c) and (f) follow from (a) as in the proof of Theorem 2. Part (d) follows from Theorem 2 and Lemma 3, and part (e) from Theorem 1 and Lemma 3.

THEOREM 4. *Let the sequences* (Q_n) *and* (k_n) *have properties* (A), (B), (C), *and* (D) *and* (E) *for* $r \geqq m$. *Suppose that* $f(x)$ *is absolutely continuous*[7] *in the set of variables* $(x_1, \cdots, x_m)$, *and in addition that when one or more of the variables* x_i *is set equal to zero* $f(x)$ *is absolutely continuous in the remaining set of variables. Let* $t(x)$ *denote a partial derivative of* $f(x)$ *in which the derivative with respect to any one variable* x_i *occurs to at most the first order, and let* $T_n(x)$ *denote the corresponding partial derivative of* $P_n(x)$. *Let* $t(x)$ *be in* L^p, $(1 \leqq p < \infty)$, *and let* A_0 *be a closed set interior to* A. *Then*

(a) $\lim_n \int_{A_0} | T_n(x) - t(x) |^p \, dx = 0$;

(b) *a subsequence of* (T_n) *converges to* t *almost everywhere on* A;

(c) *for* $E \subset A_0$ *the integrals* $\int_E | T_n(x) |^p \, dx$ *are equi-absolutely continuous*;

---

[7] That is, m-tuply absolutely continuous.

(d) *if t is bounded on A, the sequence (T_n) is uniformly bounded on A_0;*

(e) *if t is continuous at the points of A_0, the sequence (T_n) converges to t uniformly on A_0.*

(f) *the functions $P_n(x)$ are equi-absolutely continuous in the set of variables $(x_1, \cdots, x_m)$ on A_0.*

We shall carry through the proof for the case of two independent variables. The method of extending the proof to other cases and of obtaining various modifications of Theorem 4 will then be apparent. The hypothesis on the function $f(x_1, x_2)$ implies that it may be represented in the form

$$f(x_1, x_2) = f(0, 0) + \int_0^{x_1} f_{x_1}(x_1, 0)\, dx_1 + \int_0^{x_2} f_{x_2}(0, x_2)\, dx_2 + \int_0^{x_2}\int_0^{x_1} f_{x_2 x_1}\, dx_2\, dx_1.$$

It is easily seen that the hypotheses of Theorem 3 are verified, so that the desired properties of P_{nx_1} and P_{nx_2} follow from that theorem. Now the hypotheses of Lemma 3 are verified with f replaced by f_{x_2}, so that

$$\lim_n |P_n(x; f_{x_2 x_1}) - P_{nx_1}(x; f_{x_2})| = 0 \tag{3:4}$$

uniformly on A_0. Also the hypotheses of Lemma 3 are verified with Q_n replaced by $R_n = Q_{nx_1}$. Now

$$P_{nx_1}(x; f_{x_2}) = k_n \int_A f_{x_2}(z) R_n(z - x)\, dz$$

so that by Lemma 3,

$$\lim_n |P_{nx_1}(x; f_{x_2}) - P_{nx_1 x_2}(x; f)| = 0 \tag{3:5}$$

uniformly on A_0. Combining (3:4) and (3:5) we have

$$\lim_n |P_n(x; f_{x_2 x_1}) - P_{nx_1 x_2}(x; f)| = 0$$

uniformly on A_0. From this and Theorems 2 and 1 the conclusions of the theorem for $t(x) = f_{x_2 x_1}$ then follow as in the proof of Theorem 3. The final conclusion (f) follows at once from (c) with $p = 1$ and $T_n(x) = P_{nx_1 x_2}$.

4. Iteration of the approximating transformations. In the preceding discussion we have considered the linear transformations $P_n(f)$ given by formula (1:1), which transform certain normed linear spaces $\mathcal{F} = [f]$ into subsets of themselves. In certain cases these transformations are also equi-continuous, that is, there exists a constant M such that

$$\| P_n(f) \| \leq M \| f \| \tag{4:1}$$

for every n and every f in $\mathcal{F}$. They may also have the property that

$$\lim_{n} \| P_n(f) - f \| = 0 \tag{4:2}$$

for every f in $\mathcal{F}$. When the sequence (P_n) has the properties (4:1) and (4:2) and (K_n) is another sequence of transformations, not necessarily linear, but having at least property (4:2) it follows that

$$\lim_{q,\, n} \| P_q K_n f - f \| = 0. \tag{4:3}$$

Thus when the transformations K_n are also linear and have property (4:1), the composite transformations $K_n P_n$ have the same properties (4:1) and (4:2) as the factors have. In particular the iterates P_n^s, $(s = 2, 3, \cdots)$, have properties (4:1) and (4:2).

We now proceed to discuss some cases to which the argument of the preceding paragraph is applicable.

Case I. The space $\mathcal{F}$ consists of all functions f continuous on a closed rectangular region A_0 interior to A, and $\| f \|$ is the maximum of $| f(x) |$ on A_0. We extend the definition of f over the remainder of A as indicated by (1:2). Then by properties (A) and (B) and the formula (1:4), $\| P_n(f) \| \leqq \| f \|$, so that (4:1) holds true with $M = 1$. The condition (4:2) holds by Theorem 1. Note that before another transformation K_n of the same kind is applied the functions $P_n(x; f)$ are to be regarded at first as defined only for x in A_0, and then the same extension of definition over the remainder of A is to be used as was used for f itself. This procedure seems to be necessary in order that the preceding arguments may be validly applied. In examples (iv) and (v) it is easily seen that this precaution is unnecessary for many applications of the preceding result.

The iteration of the transformations P_n seems to be of special interest when the $P_n(f)$ are the integral means of f, that is, they are defined by the Q_n of example (iv). For in this case when the function f is of class $C^{(r)}$ on the region A_0, the transformed function $P_n(f)$ is of class $C^{(r+1)}$ on an arbitrary closed subregion A_1 interior to A_0, for n sufficiently large, and hence $P_n^s(f)$ is of class $C^{(r+s)}$ on an arbitrary closed subregion A_s interior to A_0, for n sufficiently large.

Case II. The space $\mathcal{F}$ consists of all functions f defined and of class L^p, $(1 \leqq p < \infty)$, on a fixed measurable subset A_0 of A, and

$$\| f \| = \left[\int_{A_0} | f |_p \, dx \right]^{1/p}.$$

In this case we agree to set $f = 0$ on the complement of A_0. Then by the same argument as that used at the beginning of the proof of Theorem 2, we find

$$\int_A | P_n(x) |^p \, dx \leqq k_n \int_B \int_A | f(x + v) |^p \, dx \, Q_n(v) \, dv \leqq k_n \int_B \| f \|^p Q_n(v) \, dv = \| f \|^p,$$

so that (4:1) holds, again with $M = 1$. Theorem 2 yields the property (4:2).

If we take A_0 to be the whole of A and the transformations $P_n(f)$ to be the integral means of f, the iterated transformations P_n^s yield approximating functions of class $C^{(s-1)}$ at least on an arbitrary closed subregion A_1 interior to A, for n sufficiently large.

Case III. The space $\mathcal{F}$ consists of all functions f absolutely continuous in the sense of Tonelli on a closed interval A_0 interior to A, and

$$\| f \| \equiv \underset{\substack{x \text{ in } A_0 \\ i=1, \cdots, m}}{\text{maximum}} \left\{ | f(x) |, \int_{A_0} | f_{x_i} | \, dx \right\}.$$

We extend the definition of f over A as indicated by (1:3), and let M denote a constant sufficiently large to satisfy all the conditions imposed on it. As in case I we find $| P_n(x; f) | \leqq \| f \|$, and as in the proof of Lemma 3 we obtain

$$\begin{aligned}
&\int_{A_0} | P_{nx_1}(x;f) - P_n(x;f_{x_1}) | \, dx \\
(4:4) \qquad &= k_n \int_{A_0} \left| \int_0^{a_2} \cdots \int_0^{a_m} [f(z) Q_n(z - x)]_{z_1=0}^{z_1=a_1} \, dz_m \cdots dz_2 \right| dx \\
&\leqq k_n \| f \| \int_A \int_0^{a_2} \cdots \int_0^{a_m} [Q_n(z - x) |^{z_1=a_1} + Q_n(z - x) |^{z_1=0}] \, dz_m \cdots dz_2 \, dx \\
&\leqq 2 \| f \| \cdot a_2 \cdots a_m \leqq M \| f \| / 2.
\end{aligned}$$

By case II we have

$$\int_{A_0} | P_n(x; f_{x_i}) | \, dx \leqq \int_A | f_{x_i} | \, dx \leqq M \| f \| / 2,$$

and combining this inequality with (4:4) (with x_1 replaced by x_i), yields

$$\int_{A_0} | P_{nx_i}(x; f) | \, dx \leqq M \| f \|,$$

so that (4:1) holds. The condition (4:2) follows from Theorems 1 and 3. We note that in this case example (iv) must be ruled out, since it does not satisfy the hypotheses of Theorem 3.

In case the function f satisfies a Lipschitz condition, and $L(f)$ denotes the minimum Lipschitz constant for f on the interval A_0, it follows from the proofs of Lemma 3 and Theorem 3 that there is a constant M such that

$$(4:5) \qquad L(P_n(f)) \leqq M[L(f) + \max | f(x) |].$$

If (K_n) is another sequence of transformations having property (4:5) and such that $| K_n(x; f) | \leqq M \cdot \max | f(x) |$, then

$$L(P_q K_n f) \leqq M^2 [L(f) + 2 \max | f(x) |],$$

so that the double sequence of functions $P_q K_n f$ satisfies for each f a uniform Lipschitz condition.

5. Approximating functions vanishing where f vanishes.[8] In this section we suppose that the function f is continuous on the closed interval A_0 interior to A, and vanishes on the closed set A_1 in A_0. We may agree to extend the domain of definition of f to be the whole of A by the method indicated in (1:2) or by that indicated in (1:3). Consider the sequence of transformations K_n defined as follows:

$$\begin{aligned} K_n(x;f) &= f(x) - 1/n \text{ where } f(x) \geqq 1/n, \\ &= f(x) + 1/n \text{ where } f(x) \leqq -1/n, \\ &= 0 \qquad\qquad \text{where } -1/n \leqq f(x) \leqq 1/n. \end{aligned}$$

Let the transformations P_n correspond to the Q_n of example (v). Then for each n there is an integer q_n sufficiently great so that the function $P_{q_n}K_nf$ vanishes on a neighborhood of the set A_1, and is of class $C^{(r)}$ on A.

In case I of Section 4 we find that the K_n satisfy (4:2) so that

$$\lim_{q,n} \| P_qK_nf - f \| = 0, \tag{5:1}$$

$$\lim_{n} \| P_{q_n}K_nf - f \| = 0. \tag{5:2}$$

In case III, where the function f is supposed to be absolutely continuous in the sense of Tonelli, we consider first the case when $f \geqq 0$. Then each partial derivative f_{x_i} vanishes wherever it exists on the set where f vanishes. Thus $K_{nx_i} = f_{x_i}$ almost everywhere on the set where $f = 0$ and almost everywhere on the set where $f > 1/n$, and $| K_{nx_i} | \leqq | f_{x_i} |$, $\lim_n K_{nx_i} = f_{x_i}$ almost everywhere. Since each function absolutely continuous in the sense of Tonelli is representable in a standard way as the difference of two non-negative functions having the same property, we find

$$\lim_n \int_{A_0} | K_{nx_i}(x;f) - f_{x_i}(x) | \, dx = 0,$$

so that (4:2) holds in this case also. Consequently we obtain (5:1) and (5:2) for this case, by the results of Section 4.

When the function f satisfies a Lipschitz condition, the transformed functions K_nf obviously satisfy the same Lipschitz condition, and hence by the last paragraph of Section 4 the functions P_qK_nf satisfy a uniform Lipschitz condition.

The results of this section show that the following theorem is valid.

THEOREM 5. *Let $f(x)$ be continuous on a bounded closed set A_0, and vanish on a set $A_1 \subset A_0$. Then there exists a sequence of functions $\varphi_n(x)$, each of class $C^{(r)}$ on the whole space and vanishing on a neighborhood of the set A_1, such that $\lim_n \varphi_n(x) = f(x)$ uniformly on A_0. In case the set A_0 is an interval and the*

[8] Cf. Reid [12], p. 859.

function f is absolutely continuous in the sense of Tonelli, the sequence (φ_n) *may be required to satisfy also the condition that*

$$\lim_n \int_{A_0} | \phi_{nx_i} - f_{x_i} | \, dx = 0.$$

When the function f satisfies a Lipschitz condition on the set A_0, the functions φ_n may be required to satisfy also a uniform Lipschitz condition on A_0.

Note that the last statement is valid when the set A_0 is an arbitrary bounded set, since a Lipschitz function may have its domain of definition extended over the whole of space without losing that property.

THE UNIVERSITY OF CHICAGO.

REFERENCES

(1) Landau, *Über die approximation einer stetigen funktion durch eine ganze rationale funktion*, Rendiconti del Circolo Matematico di Palermo, vol. 25 (1908), pp. 337–345.

(2) la Vallée Poussin, *Cours d'Analyse*, 2nd Ed., Tome 2 (1912), pp. 126–37.

(3) Tonelli, *Sulla rappresentazione analitica delle funzioni di più variabili reali*, Rendiconti del Circolo Matematico di Palermo, Vol. 29 (1910), pp. 1–36.

(4) Tonelli, *Sopra alcuni polinomi approssimativi*, Annali di Matematica (III), vol. 25 (1916), pp. 275-316.

(5) Cinquini, *Sull'approssimazione delle funzioni di due variabili*, Annali di Matematica, (IV) vol. 11 (1933), pp. 295-323.

(6) Cinquini, *Su una proprietà dei polinomi di Stieltjes*, Rendiconti del Circolo Matematico di Palermo, vol. 58 (1934), pp. 57–72.

(7) Cinquini, *Sul problema dell' approssimazione delle funzioni*, Annali della R. Scuola Normale Superiore, Pisa, (II) vol. 4 (1935), pp. 85-103.

(8) Hobson, *The theory of functions of real variables*, 2nd Edition, (vol. I, 1921; vol. II. 1926).

(9) Saks, *Theory of the Integral* (1937).

(10) McShane, *Extension of range of functions*, Bulletin of the American Mathematical Society, vol. 40 (1934), pp. 837–842.

(11) Whitney, *Analytic extensions of differentiable functions defined in closed sets*, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 63-89.

(12) Reid, *The Jacobi condition for the double integral problem of the calculus of variations*, Duke Mathematical Journal, vol. 5 (1939), pp. 856-870

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

TAUBERIAN CONDITIONS[1]

BY RALPH PALMER AGNEW

(Received March 30, 1940)

1. Introduction. It is the object of this paper to introduce several new Tauberian classes and to discuss their relations to each other and to various Tauberian classes which have been previously considered.

As the title indicates, our interest lies in Tauberian conditions themselves rather than in Tauberian theorems involving them. In order that our discussion of Tauberian theorems may be as free as possible from complications which would draw attention from the Tauberian conditions, we consider in this paper only Tauberian theorems for the C_1 (Cesàro of order 1, or arithmetic mean) method of summability. That the fundamental Fourier transform method of Wiener[2] can be applied to give general Tauberian theorems for the classes T, T^*, T'_p, and T' which we shall define is shown by a paper of H. R. Pitt to which we refer in §10. It is probable that the methods of Wiener and Pitt apply also to the more general classes T''_B of §11. Each of the classic C_1 Tauberian theorems to which we refer in §2 is a corollary of Theorem 9.1.

2. Classic Tauberian theorems. Let $u_1 + u_2 + \cdots$ be a series (convergent or divergent) of complex terms, and let

$$s_n = \sum_{k=1}^{n} u_k; \qquad \sigma_n = \frac{1}{n}\sum_{k=1}^{n} s_k \qquad\qquad n = 1,2, \cdots$$

be respectively the sequence of partial sums and the C_1 transform of $\sum u_n$. An elementary *Abelian theorem* states that if $s_n \to s$, then $\sigma_n \to s$; in other words, that C_1 is a regular method of summability. That $\sigma_n \to \sigma$ does not imply $s_n \to \sigma$ is illustrated by the series $\sum (-1)^{n+1}$ for which $\sigma_n \to \frac{1}{2}$. If, however, $\sigma_n \to \sigma$ and the terms of $\sum u_n$ satisfy certain auxiliary *Tauberian conditions*, then $s_n \to \sigma$. The following, in which $\tilde{T}$ denotes a class of series, is a typical Tauberian theorem.

2.1 *If* $\sigma_n \to \sigma$ *and* $\sum u_n \in \tilde{T}$, *then* $s_n \to \sigma$.

The essential part of the conclusion is that s_n converges; for if it is known that s_n converges, then regularity of C_1 implies that $s_n \to \sigma$.

If $\tilde{T}$ is the class T_C of convergent series, then 2.1 is true but trivial since the class T is too small to make 2.1 significant. If $\tilde{T}$ is the union of the class T_C and the class T_D of all divergent (non-convergent) series not summable C_1, then again 2.1 is obviously true; but again the theorem is trivial since the defini-

[1] Presented to the American Mathematical Society, February 24, 1940.

[2] N. Wiener, Tauberian Theorems, Annals of Math., vol. 33 (1932), pp. 1–100.

tion of $\bar{T}$ requires that one know whether s_n converges *before* one can decide whether $\sum u_n$ satisfies the hypothesis of 2.1. If 2.1 is to be useful, there must naturally be some series for which it is easier to determine that $\sum u_n \epsilon \bar{T}$ than it is to determine whether $\sum u_n$ converges. For application to a given series $\sum u_n$ belonging to some Tauberian class $\bar{T}$, the best theorems are those involving $\bar{T}$ for which it is easiest to show that $\sum u_n \epsilon \bar{T}$. These considerations indicate that there probably is no single "best" Tauberian class which will eventually supplant all others.

A simple Tauberian theorem states that if $\sigma_n \to \sigma$ and $\sum u_n$ satisfies the condition $nu_n \to 0$ of Tauber[3] then $s_n \to \sigma$. A more general Tauberian theorem asserts that if $\sigma_n \to \sigma$ and $\sum u_n$ satisfies the condition $n \mid u_n \mid < K$ of Hardy[4] then $s_n \to \sigma$. A unilateral Tauberian theorem states that if $\sigma_n \to \sigma$ and $\sum u_n$ is a real series satisfying the condition $nu_n < K$ (or the condition $nu_n > -K$) of Landau[5] then $s_n \to \sigma$. The last theorem was extended by Lukacs[6] to cover complex series, by the next theorem in which S' denotes the set of points in the complex plane which lie in a sector with vertical angle less than π. If $\sigma_n \to \sigma$ and $nu_n \epsilon S'$, then $s_n \to \sigma$. The conditions $nu_n \to 0$, $n \mid u_n \mid < K$, $nu_n < K$, and $nu_n \epsilon S'$ (as well as some conditions of Schmidt which we discuss in §7) are known as *order conditions*. The Tauberian condition of the next theorem is known as a *gap condition*.[7] If $\sigma_n \to \sigma$ and $u_n = 0$ when $n \neq n_1, n_2, \cdots$, where $n_1 < n_2 < \cdots$ is a sequence of indices for which $\liminf_{p\to\infty} n_{p+1}/n_p > 1$, then $s_n \to \sigma$. "Gap Tauberian theorems" such as the above have been called[8] remarkable since there is no order condition on the terms u_n.

[3] A. Tauber. Ein Satz aus der Theorie der unendlichen Reihen, Monatshefte für Mathematik und Physik, vol. 8 (1897), pp. 273–277.

[4] G. H. Hardy, Theorems relating to the summability and convergence of slowly oscillating series, Proceedings of the London Mathematical Society, Series 2, vol. 8 (1909), pp. 301–320.

[5] E. Landau, Über die Bedeutung einiger neuerer Grenzwertsätze der Herren Hardy und Axer, Prace Matematyczno-Fizyczne, vol. 21 (1910), pp. 97–177. See also E. Landau, Darstellung und Begrundung einiger neuerer Ergebnisse der Funktionentheorie, 2nd Edition, Berlin, 1929.

[6] F. Lukacs, Bemerkung zu einem Konvergenzsatze des Herrn Landau, Arch. d. Math. u. Phys. (3) 23 (1915), pp. 367–378. For further results and references, see L. S. Bosqnquet and M. L. Cartwright, Mathematische Zeitschrift, vol. 37 (1933), pp. 416–423; and N. Wiener, ibid., vol. 36 (1933), pp. 787–789.

[7] See Hardy and Littlewood, A further note on the converse of Abel's theorem, Proceedings of the London Mathematical Society, Series 2, vol. 25 (1926), pp. 219–236. This fundamental paper removes order conditions from the hypotheses of Tauberian theorems due to Landau (1907) and Littlewood (1910). For a recent and brief proof and discussion of gap theorems, see A. E. Ingham, On the high-indices theorem of Hardy and Littlewood, Quarterly Journal of Math., Oxford Series, vol. 8 (1937), pp. 1–7.

[8] For a recent example, see N. Levinson, General Gap Tauberian Theorems I, Proceedings of the London Mathematical Society, Series 2, vol. 44 (1938), pp. 289–306, p. 289.

3. The class T. Let K be a positive constant, and let θ be an angle for which $0 < \theta < \pi/2$. Corresponding to each real angle ψ, let $S(\psi) \equiv S(K, \theta, \psi)$ denote the "sector with vertical angle $< \pi$" consisting of all points z of the complex plane representable in the form

$$z = -Ke^{i\psi} + \rho e^{i(\psi+\varphi)}$$

where $\rho \geqq 0$ and $-\theta \leqq \varphi \leqq \theta$. The sector $S(\psi)$ has its vertex at the point $-Ke^{i\psi}$; the angle at the vertex of the sector is 2θ; the half-line bisecting the sector passes through the origin and makes an angle ψ with the positive real axis; and each sector $S(\psi)$ may be obtained by rotating the particular sector $S(0)$ about the origin. The sectors $S(\psi)$ are special sectors in that the bisectors pass through the origin; however if S' is any sector with vertical angle $< \pi$, then it is easy to determine K, θ, and ψ so that S' is a subset of $S(K, \theta, \psi)$.

Definition 3.1. *A series $\sum u_n$ will be said to belong to class T if K, θ, λ, and $\psi_1, \psi_2, \psi_3, \cdots$ exist such that $K > 0$, $0 < \theta < \pi/2$, $\lambda > 1$, and, for each sufficiently great index k,*

$$nu_n \in S(\psi_k) \equiv S(K, \theta, \psi_k) \qquad k \leqq n < \lambda k. \tag{3.11}$$

The Tauberian condition $nu_n \in S'$ of §2 implies that there is a *fixed* sector S' which contains all elements of the sequence nu_n; hence the condition $nu_n \in S'$ may be termed a *fixed-sector condition*. On the other hand the Tauberian condition $\sum u_n \in T$ allows *different* sectors S to contain different aggregates of consecutive elements of the sequence nu_n; hence the condition $\sum u_n \in T$ may be termed a *variable-sector condition*.

A real series $\sum u_n$ satisfies the fixed-sector condition $nu_n \in S'$ if and only if it satisfies at least one of the two unilateral conditions $nu_n < K$ and $nu_n > -K$. The condition $\sum u_n \in T$ does not imply either of these unilateral conditions. In fact it is easy to show that a real series $\sum u_n$ is in class T if and only if $K > 0$ and $\lambda > 1$ exist such that for each sufficiently great k, say $k > k_0$, either

$$nu_n < K \qquad k \leqq n < \lambda k \tag{3.12}$$

or

$$nu_n > -K \qquad k \leqq n < \lambda k. \tag{3.13}$$

The essential point here is that (3.12) may hold for a certain set of values of $k > k_0$ while (3.13) holds for the remaining values of $k > k_0$.

4. Some special series in class T. The next two theorems show that if $\sum u_n$ satisfies one of the Tauberian conditions of §2, then $\sum u_n \in T$.

Theorem 4.1. *If $nu_n \in S'$, where S' is some sector of the complex plane with vertical angle $< \pi$, then $\sum u_n \in T$.*

This is obvious since we can choose a sector $S(K, \theta, \psi)$ such that $S' \subset S(K, \theta, \psi)$ and then take $\psi_k = \psi$ for each $k = 1, 2, \cdots$ to show that $\sum u_n \in T$. It

follows that if $\sum u_n$ satisfies one of the other order conditions of §2, then $\sum u_n \in T$.

THEOREM 4.2. *If $n_1 < n_2 < \cdots$ is a gap sequence such that*

$$\liminf_{p\to\infty} n_{p+1}/n_p > 1,$$

and $u_n = 0$ when $n \neq n_1, n_2, \cdots$, then $\sum u_n \in T$.

Choose $\lambda > 1$ such that $\liminf n_{p+1}/n_p > \lambda$, and then choose an index P such that

$$n_{p+1} > \lambda n_p \qquad p \geqq P.$$

For each $k \geqq n_p$ there is at most one index n in the interval $k \leqq n < \lambda k$ for which $u_n \neq 0$. We can assume that $u_n \neq 0$ for an infinite set of n for otherwise certainly $\sum u_n \in T$. For each $k = 1, 2, \cdots$ let ψ_k be the amplitude φ_n of $u_n = \rho_n \exp i\varphi_n$ where $\rho_n > 0$, $-\pi < \varphi_n \leqq \pi$, and n is the least index such that $n \geqq k$ and $u_n \neq 0$. Then, however $K > 0$ and $0 < \theta < \pi/2$ are determined, the relation

$$nu_n \in S(K, \theta, \psi_k) \qquad k \leqq m < \lambda k$$

holds when $k \geqq k_P$ and accordingly $\sum u_n \in T$. The series satisfying the hypothesis of Theorem 4.2 are gap series with no condition whatever on u_n when n has one of the values $n_1, n_2, n_3, \cdots$.

It is possible to generalize Theorem 4.2 in various ways; for example, the following theorem is an immediate consequence of Theorems 4.2 and 6.1.

THEOREM 4.3. *If $n_1 < n_2 < \cdots$ is a gap sequence as in Theorem 4.2 and*

$$n \mid u_n \mid < K \qquad n \neq n_1, n_2, \cdots$$

then $\sum u_n \in T$.

Each series $\sum u_n$ satisfying the hypothesis of Theorem 4.3 is representable in the form $\sum (u_n + v_n)$ where each one of the series $\sum u_n$ and $\sum v_n$ satisfies a classic Tauberian condition of §2. That T contains important classes of series not so representable is illustrated by the fact that T contains each ordinary Dirichlet series with non-negative coefficients.

THEOREM 4.4. *If $a_n \geqq 0$ then $\sum a_n n^{-z} \in T$ for each complex z.*

Setting $u_n = a_n n^{-z}$ and $z = x + iy$ where x and y are real, we find that $nu_n = A_n \exp i\varphi_n$ where $A_n = na_n n^{-x} \geqq 0$ and $\varphi_n = -y \log n$. If $0 < \theta < \pi/2$, and $\lambda > 1$ is chosen such that $\mid y \mid \log \lambda < \theta$, then

$$\mid \varphi_n - \varphi_k \mid \leqq \mid y \mid \log \lambda < \theta \qquad k \leqq n < \lambda k$$

and it follows that $\sum u_n \in T$. In case the argument z in the Dirichlet series $\sum a_n n^{-z}$ is pure imaginary say $z = iy$, the hypothesis $a_n \geqq 0$ can be relaxed (see Theorem 6.2) to give

THEOREM 4.5. *If $\sum a_n \in T$, then $\sum a_n n^{-iy} \in T$ for each real y.*

In particular, if a_n is real and $a_n < K$ (or $a_n > -K$), then $\sum a_n n^{-z} \in T$ for each z on the line $\Re z = 1$.

It would be possible to mention other interesting subclasses of T; we have merely given enough to indicate that Tauberian theorems which are proved for the class T have diverse applications.

5. Characterizations of T. The following two quite trivial theorems give characterizations of T which are useful in proofs of Tauberian theorems involving T.

THEOREM 5.1. *In order that $\sum u_n \in T$, it is necessary and sufficient that $K > 0$, $0 < \theta < \pi/2$, $\alpha < 1 < \beta$, and $\psi_1, \psi_2, \cdots$ exist such that for each sufficiently great index k*

$$(5.11) \qquad nu_n \in S(\psi_k) \equiv S(K, \theta, \psi_k) \qquad \alpha k < n < \beta k.$$

Sufficiency is obvious. To prove necessity, let $\sum u_n \in T$ and choose $K > 0$, $0 < \theta < \pi/2$, $\lambda > 1$, and $\psi_1', \psi_2', \cdots$ such that for each sufficiently great k

$$(5.12) \qquad nu_n \in S(K, \theta, \psi_k') \qquad k \leqq n < \lambda k.$$

If in (5.12) we replace k by $k' = [k\lambda^{-1/2}]$ and choose α and β such that

$$(5.13) \qquad \lambda^{-1/2} < \alpha < 1 < \beta < \lambda^{1/2}$$

we find that, for k sufficiently great,

$$(5.14) \qquad nu_n \in S(K, \theta, \psi_{k'}') \qquad \alpha k < n < \beta k.$$

Thus (5.11) holds when ψ_k is $\psi_{k'}'$ and the theorem is proved.

THEOREM 5.2. *In order that $\sum u_n \in T$, it is necessary and sufficient that $K > 0$, $0 < \theta < \pi/2$, $\mu < 1$, and $\psi_1, \psi_2, \cdots$ exist such that for each sufficiently great index k*

$$(5.21) \qquad nu_n \in S(\psi_k) \equiv S(K, \theta, \psi_k) \qquad \mu k < n \leqq k.$$

Necessity is implied by Theorem 5.1. Proof that (5.21) implies (5.11), and hence $\sum u_n \in T$, differs very little from our proof that (5.12) implies (5.11).

6. Properties of T. The definition 3.1 of T obviously implies that if $\sum u_n \in T$ and $u_n' = u_n$ for all sufficiently great n, then $\sum u_n' \in T$. In particular if $u_1 + u_2 + u_3 + \cdots \in T$, then $(u_1 - A) + u_2 + u_3 + \cdots \in T$ for each complex number A.

It is easy to show that if $\sum \mathcal{R} u_n \in T$ and $\sum \mathcal{I} u_n \in T$, then $\sum u_n \in T$; that is, that $\sum u_n$ belongs to class T if the series of real parts and the series of pure imaginary parts belong to the class. However simple examples show that $\sum u_n \in T$ implies neither $\sum \mathcal{R} u_n \in T$ nor $\sum \mathcal{I} u_n \in T$. The class T is not linear. For example, if $u_n' = (-1)^n + 1$ and $u_n'' = (-1)^n - 1$, then $\sum u_n' \in T$ and $\sum u_n'' \in T$; but $\sum (u_n' + u_n'') \in T$ fails.

THEOREM 6.1. *If $\sum u_n \in T$ and $n \,|\, v_n \,| < K_1$, then $\sum (u_n + v_n) \in T$.*

Choose $K > 0$, $0 < \theta < \pi/2$, $\lambda > 1$, and $\psi_1, \psi_2, \cdots$, and an index k_0 such that when $k > k_0$

$$nu_n \in S(K, \theta, \psi_k) \qquad k \leqq n < \lambda k. \tag{6.11}$$

Elementary computation shows that

$$n(u_n + v_n) \in S(K + K_1 \csc \theta, \theta, \psi_k) \qquad k \leqq n < \lambda k, \tag{6.12}$$

and hence that $\sum (u_n + v_n) \in T$.

THEOREM 6.2. *If $\sum u_n \in T$ and $\beta_1, \beta_2, \cdots$ is a real sequence such that*

$$\lim_{\lambda \to 1+} \limsup_{k \to \infty} \max_{k \leqq n < \lambda k} |\beta_n - \beta_k| = 0 \tag{6.21}$$

then $\sum u_n \exp i\beta_n \in T$.

Obtaining (6.11) as above, we have, when $k > k_0$ and $k \leqq n < \lambda k$

$$nu_n = -Ke^{i\psi_k} + \rho_n e^{i(\psi_k + \varphi_n)} \tag{6.22}$$

where $\rho_n \geqq 0$ and $|\varphi_n| \leqq \theta < \pi/2$. Choose $\theta' > 0$ such that $\theta'' \equiv \theta + \theta' < \pi/2$. Let λ be decreased and let k_0 be increased if necessary so that, when $k > k_0$ and $k \leqq n < \lambda k$, (6.22) holds and also $|\beta_n - \beta_k| < \theta'$. If we set $\psi_k' = \psi_k + \beta_k$ and $\varphi_n' = \varphi_n + \beta_n - \beta_k$ (our notation does not take into account the fact that φ_n' as well as ρ_n and φ_n depend on k) we obtain

$$nu_n e^{i\beta_n} = -Ke^{i(\psi_k + \beta_n)} + \rho_n e^{i(\psi_k' + \varphi_n')} \tag{6.23}$$

where $\rho_n \geqq 0$ and $|\varphi_n'| \leqq |\varphi_n| + |\beta_n - \beta_k| \leqq \theta''$. It can be shown that $nu_n \exp i\beta_n \in S(K \csc \theta'', \theta'', \psi_k')$ when $k > k_0$ and $k \leqq n < \lambda k$ and Theorem 6.2 is proved.

THEOREM 6.3. *If $\sum u_n \in T$ and $a_1, a_2, \cdots$ is a bounded sequence of real non-negative constants, then $\sum a_n u_n \in T$.*

Obtaining (6.22) as above, we find that

$$na_n u_n = -Ka_n e^{i\psi_n} + a_n \rho_n e^{i(\psi_k - \varphi_n)}.$$

If K' is the least upper bound of the non-negative numbers Ka_n, then

$$na_n u_n = -K' e^{i\psi_n} + \rho_n' e^{i(\psi_k - \varphi_n')}$$

where $\rho_n' \geqq 0$ and $-\theta \leqq \varphi_n' \leqq \theta$; hence $\sum a_n u_n \in T$.

Theorem 6.3 can be used to show that if $\sum u_n \in T$, then each series obtained by adjoining or removing a finite set of terms at the beginning of the series is also in class T; if q is a positive integer and one of the two series

$$u_1 + u_2 + u_3 + \cdots$$

$$u_q + u_{q+1} + u_{q+2} + \cdots$$

is in class T, so also is the other

7. The class T^*. The developments of this section are largely extensions of the ideas of "langsam oszillierend" and "langsam abfallende" sequences introduced by Schmidt.[9]

It is well known that if $\sum u_n$ satisfies the Tauberian condition $n \mid u_n \mid < K$ then the simple estimate,

$$| s_q - s_p | \leqq \sum_{n=p+1}^{q} | u_n | \leqq K \sum_{n=p+1}^{q} n^{-1}$$
$$\leqq K(\log q - \log p + o_k) \leqq K \log (\beta/\alpha) + o_k,$$

in which $\alpha k < p < q < \beta k$ and $o_k \to 0$ as $k \to \infty$, shows that the partial sums s_n of $\sum u_n$ satisfy the Tauberian condition

$$\lim_{\alpha \to 1-, \beta \to 1+} \limsup_{k \to \infty} \operatorname*{maximum}_{\alpha k < p < q < \beta k} | s_q - s_p | = 0. \tag{7.01}$$

Since

$$\operatorname*{maximum}_{k \leqq n < \beta k} | s_n - s_k | \leqq \operatorname*{maximum}_{\alpha k < p < q < \beta k} | s_q - s_p |$$

when $\alpha < 1 < \beta$, the condition (7.01) implies that

$$\lim_{\lambda \to 1+} \limsup_{k \to \infty} \operatorname*{maximum}_{k \leqq n < \lambda k} | s_n - s_k | = 0; \tag{7.011}$$

and it is easy to show that (7.011) implies (7.01).

It is likewise well known that if $\sum u_n$ is a real series satisfying the unilateral condition $nu_n > -K$, then a similar computation shows that

$$\lim_{\alpha \to 1-, \beta \to 1+} \liminf_{k \to \infty} \operatorname*{minimum}_{\alpha k < p < q < \beta k} (s_q - s_p) \geqq 0. \tag{7.02}$$

If (7.02) holds, then obviously

$$\lim_{\lambda \to 1+} \liminf_{k \to \infty} \operatorname*{minimum}_{k \leqq n < \lambda k} (s_n - s_k) \geqq 0. \tag{7.021}$$

If (7.021) holds, then corresponding to each $\epsilon > 0$ there exists $\lambda_0 > 1$ such that when $1 < \lambda < \lambda_0$ there exists $k_0 = k_0(\lambda)$ such that $(s_n - s_k) > -\epsilon$ when $k > k_0$ and $k \leqq n < \lambda k$. If $1 < \lambda < \lambda_0$ and $\lambda^{-1/2} < \alpha < 1 < \beta < \lambda^{1/2}$ and $\alpha k < p \leqq r \leqq q < \beta k$ then, for k sufficiently great, $(s_r - s_p) > -\epsilon$ and $(s_q - s_r) > -\epsilon$ so that $(s_q - s_p) > -2\epsilon$; and hence we can obtain (7.02). Thus the Tauberian conditions (7.02) and (7.021) are equivalent. Schmidt (loc. cit., p. 136) has called a real series "langsam abfallend" when its partial sums s_n have the following property: If $q = q(p)$ is a sequence of indices such that $q \geqq p$ for each $p = 1, 2, \cdots$ and $q/p \to 1$ as $p \to \infty$, then $\liminf_{p \to \infty}(s_q - s_p) \geqq 0$. Using this definition, it is easy to show that the class of "langsam abfallend" series is identical with the class of series satisfying (7.021), and hence also identical with the class of series satisfying (7.02).

---

[9] Robert Schmidt, Über divergente Folgen und lineare Mittelbildungen, Mathematische Zeitschrift, vol. 22 (1925), pp. 89–152; p. 132 and p. 136.

In a similar manner, it can be shown that if a real series $\sum u_n$ satisfies the Tauberian condition $nu_n < K$, then $\sum u_n$ satisfies the equivalent Tauberian conditions

$$\lim_{\alpha\to 1-,\,\beta\to 1+} \limsup_{k\to\infty} \underset{\alpha k<p<q<\beta k}{\text{maximum}} (s_q - s_p) \leqq 0. \tag{7.03}$$

and

$$\lim_{\lambda\to 1+} \limsup_{k\to\infty} \underset{k\leqq n<\lambda k}{\text{maximum}} (s_n - s_k) \leqq 0. \tag{7.031}$$

That (7.02) and (7.03) are essentially different conditions is evidenced by the fact that the sequence $s_n = n$ satisfies (7.02) but not (7.03), and the sequence $s_n = -n$ satisfies (7.03) but not (7.02).

Numerous Tauberian theorems involving series satisfying such conditions as (7.01) $\cdots$ (7.031) can be found in a paper of Szasz,[10] a monograph of Karamata,[11] and references given by these authors. A double series analogue of (7.02) has been used by Knopp[12] to obtain Tauberian theorems for double series.

We proceed to show that if $\sum u_n \in T$, then the partial sums of $\sum u_n$ satisfy a condition more general than (7.01), (7.02), and (7.03). Let $\sum u_n \in T$. Then $K > 0$, $0 < \theta < \pi/2$, $\alpha_0 < 1 < \beta_0$, and $-\psi_1, -\psi_2, \cdots$ exist such that for each sufficiently great k

$$nu_n = -Ke^{-i\psi_k} + \rho_n e^{i(-\psi_k+\varphi_n)} \qquad \alpha_0 k < n < \beta_0 k \tag{7.04}$$

where $\rho_n \geqq 0$ and $|\varphi_n| \leqq \theta$. If $\alpha_0 < \alpha < 1 < \beta < \beta_0$, then for each real δ_k

$$nu_n e^{i(\psi_k+\delta_k)} = -Ke^{i\delta_k} + \rho_n e^{i(\varphi_n+\delta_k)} \tag{7.05}$$

when $\alpha k < n < \beta k$. Let

$$\Delta = \pi/2 - \theta. \tag{7.06}$$

If $|\delta_k| \leqq \Delta$, then (7.05) and the inequalities $\rho_n \geqq 0$ and $|\varphi_n + \delta_k| \leqq \theta + \Delta = \pi/2$ imply that

$$\mathcal{R} nu_n e^{i(\psi_k+\delta_k)} \geqq -K \qquad \alpha k < n < \beta k. \tag{7.07}$$

Hence, when $|\delta_k| \leqq \Delta$ and $\alpha k < p < q < \beta k$,

$$\begin{aligned} \mathcal{R}(s_q - s_p) e^{i(\psi_k+\delta_k)} &= \sum_{n=p+1}^{q} \mathcal{R} u_n e^{i(\psi_k+\delta_k)} \geqq -K \sum_{n=p+1}^{q} \frac{1}{n} \\ &\geqq -K \log(\beta/\alpha) + o_k \end{aligned} \tag{7.08}$$

---

[10] O. Szasz, Converse theorems of summability for Dirichlet's series, Trans. American Math. Soc., 39 (1936), pp. 117–130.

[11] J. Karamata, Sur les theoremes inverses des procedes de sommabilite, Hermann and Cie, Paris (1937), 47 pp.

[12] K. Knopp, Limitierungs—Umkehrsätze für Doppelfolgen, Math. Zeit., vol. 45 (1939), pp. 573–589.

where o_k denotes a quantity which, when α and β are fixed, converges to 0 as $k \to \infty$.

Setting

$$F_1(\alpha, \beta, k, \delta_k, \psi_k) = \underset{\alpha k < p < q < \beta k}{\text{mimimum}} \mathcal{R}(s_q - s_p) e^{i(\psi_k + \delta_k)} \tag{7.09}$$

we see that when $|\delta_k| \leqq \Delta$

$$F_1(\alpha, \beta, k, \delta_k, \psi_k) \geqq -K \log (\beta/\alpha) + o_k. \tag{7.10}$$

Setting

$$F_2(\alpha, \beta, k) = \max_{-\infty < \psi_k < \infty} \min_{|\delta_k| \leqq \Delta} F_1(\alpha, \beta, k, \delta_k, \psi_k) \tag{7.11}$$

we see that

$$F_2(\alpha, \beta, k) \geqq -K \log (\beta/\alpha) + o_k; \tag{7.12}$$

hence

$$\liminf_{k \to \infty} F_2(\alpha, \beta, k) \geqq -K \log (\beta/\alpha) \tag{7.13}$$

and finally, since the left member of (7.13) is monotone increasing as $\alpha \to 1-$, $\beta \to 1+$,

$$\lim_{\alpha \to 1-, \beta \to 1+} \liminf_{k \to \infty} F_2(\alpha, \beta, k) \geqq 0. \tag{7.14}$$

DEFINITION 7.2. *Let T^* denote the class of series $\sum u_n$ whose partial sums satisfy the condition: a constant Δ exists such that*[13] $0 < \Delta < \pi/2$ *and*

$$\begin{aligned} &\lim_{\alpha \to 1-, \beta \to 1+} \liminf_{k \to \infty} \underset{-\infty < \psi_k < \infty}{\text{maximum}} \\ &\underset{|\delta_k| \leqq \Delta}{\text{minimum}} \underset{\alpha k < p < q < \beta k}{\text{minimum}} \mathcal{R}(s_q - s_p) e^{i(\psi_k + \delta_k)} \geqq 0. \end{aligned} \tag{7.21}$$

Since (7.21) is merely the result of substituting for $F_2(\alpha, \beta, k)$ in (7.14), and we have shown that (7.14) is implied by the hypothesis that $\sum u_n \in T$, we have proved

THEOREM 7.3. *If $\sum u_n \in T$, then $\sum u_n \in T^*$.*

The condition (7.21) would be unchanged in meaning if the subscript k were removed from ψ_k and δ_k; we retain the subscript because it serves to emphasize the meaning of (7.21). It is the "variable sector" feature of the class T which accounts for the presence of the ψ's in (7.21), and it is the presence of the ψ's in (7.21) which gives the class T its generality. It should of course be observed that, in the parade of limits and extrema in (7.12), the operation of taking the maximum over $-\infty < \psi_k < \infty$ is the only one which strives to make the inequality sign in (7.12) run the right way. The condition that $\sum u_n \in T^*$ (as well as the condition $\sum u_n \in T$) may be stated roughly as follows: in the

[13] The left member of (7.21) is a monotone increasing function of Δ; hence if (7.21) holds for some $\Delta > 0$, we can suppose $0 < \Delta < \pi/2$.

infinite succession of "jumps" in the complex plane from s_1 to s_2 to s_3 to $\cdots$, jumps which are advanced in the sequence and too close together must not be both too large and in directions too nearly opposite.

8. Properties of T^*. It is easy to show that if $\sum u_n$ satisfies one of the conditions (7.01), (7.02), and (7.03), then $\sum u_n \in T^*$. That T^* includes series $\sum u_n$ for which (7.01), (7.02), and (7.03) all fail is illustrated by large classes of series and in particular by many real gap series. That T^* contains series not in T is illustrated by the fact that T^* contains each convergent series while T does not. For example, if $s_n = (-1)^n/(n+1)^{1/2}$, then $\sum u_n$ is not in the class T. It is apparent from (7.21) that if $u_1 + u_2 + \cdots$ is a series in T^*, then $(u_1 - A) + u_2 + \cdots$ is a series in T^* for each complex constant A.

It is possible to characterize T^* by conditions similar to but different from (7.21). If we replace ψ_k by $\psi_k + \pi$ in (7.21) and remove the factor $e^{i\pi} = -1$, we see that $\sum u_n \in T^*$ if and only if $\Delta > 0$ exists such that

$$\begin{aligned}(8.01)\qquad &\lim_{\alpha\to 1-,\,\beta\to 1+}\ \limsup_{k\to\infty}\ \underset{-\infty<\psi_k<\infty}{\text{minimum}}\\ &\underset{|\delta_k|\leqq\Delta}{\text{maximum}}\ \underset{\alpha k<p<q<\beta k}{\text{maximum}}\ \mathcal{R}(s_q - s_p)e^{i(\psi_k+\delta_k)} \leqq 0.\end{aligned}$$

It thus appears that, while (7.02) and (7.03) are essentially different conditions, (7.21) and (8.01) differ only in appearance.

THEOREM 8.1. *In order that $\sum u_n \in T^*$, it is necessary and sufficient that $\Delta > 0$ exist such that*

$$\begin{aligned}(8.11)\qquad &\lim_{\lambda\to 1+}\ \liminf_{k\to\infty}\ \underset{-\infty<\psi_k<\infty}{\text{maximum}}\\ &\underset{|\delta_k|<\Delta}{\text{minimum}}\ \underset{k\leqq n<\lambda k}{\text{minimum}}\ \mathcal{R}(s_n - s_k)e^{i(\psi_k+\delta_k)} \geqq 0.\end{aligned}$$

Necessity is a consequence of the fact that if $\alpha < 1 < \beta = \lambda$, and $R_{n,k}$ is a real double sequence, then

$$(8.12)\qquad \underset{k\leqq n<\lambda k}{\text{minimum}}\ R_{n,k} \geqq \underset{\alpha k<p<q<\beta k}{\text{minimum}}\ R_{p,q}.$$

To prove sufficiency, suppose (7.21) fails. Then, when $\Delta > 0$, a constant $H = H(\Delta) > 0$ exists such that the left member of (7.21) is less than $-H$. Hence $\alpha_0 < 1 < \beta_0$ exist such that, when α and β are fixed numbers satisfying the inequality $\alpha_0 < \alpha < 1 < \beta < \beta_0$, the inequality

$$(8.13)\qquad \underset{-\infty<\psi_k<\infty}{\text{maximum}}\ \underset{|\delta_k|\leqq\Delta}{\text{minimum}}\ \underset{\alpha k<p<q<\beta k}{\text{minimum}}\ \mathcal{R}(s_q - s_p)e^{i(\psi_k+\delta_k)} < -H$$

holds for an infinite set of values of the index k, say $k_1 < k_2 < \cdots$. When k has a fixed one of the values $k_1, k_2, \cdots$ there exist indices p_k and q_k such that

$$(8.14)\qquad \alpha k < p_k < q_k < \beta k$$

and

$$(8.15)\qquad \underset{-\infty<\psi<\infty}{\text{maximum}}\ \underset{|\delta|<\Delta}{\text{minimum}}\ \mathcal{R}(s_{q_k} - s_{p_k})e^{i(\psi+\delta)} < -H.$$

Since (8.14) implies that $p_k \leqq q_k < \lambda p_k$ where $\lambda = \beta/\alpha > 1$, we can show that (8.15) contradicts (8.11) and Theorem 8.1 is proved.

The condition (8.11) is more convenient than (7.21) when one wishes to show that $\sum u_n \epsilon T^*$; but the condition (7.21) which we have featured in the definition of T^* is more convenient than (8.11) when the condition $\sum u_n \epsilon T^*$ is used as a part of the hypothesis of a Tauberian theorem.

Our definition and discussion of T^* naturally applies to series $\sum u_n$ whose terms are real. Since much of the existing Tauberian theory applies only to series with real terms (though in many cases there is an immediate application to series of complex terms of which the real and imaginary parts separately satisfy the Tauberian conditions involved) it may be of interest to see that the condition (7.21) can be thrown into a form somewhat simpler when $\sum u_n$ is real. Using the definition 7.2 of T^*, we can prove

THEOREM 8.2. *If the terms of $\sum u_n$ are real, then $\sum u_n \epsilon T^*$ if and only if*

$$(8.21) \qquad \lim_{\alpha \to 1-, \beta \to 1+} \limsup_{k \to \infty} \max_{j_k = 1,2} \operatorname{minimum}_{\alpha k < p < q < \beta k} (-1)^{j_k}(s_q - s_p) \geqq 0.$$

It can also be shown that (8.21) may be replaced in Theorem 8.2 by the condition

$$(8.22) \qquad \lim_{\lambda \to 1+} \limsup_{k \to \infty} \max_{j_k = 1,2} \operatorname{minimum}_{k \leqq n < \lambda k} (-1)^{j_k}(s_n - s_k) \geqq 0.$$

When the conditions (8.21) and (8.22) are made more restrictive by removal of the expressions involving j, they reduce to the characterizations (7.02) and (7.021) of the "langsam abfallend" series of Schmidt.

9. Tauberian theorems for C_1 summability. We now prove the two following theorems of which the first is a Tauberian convergence theorem and the second is a Tauberian oscillation theorem.

THEOREM 9.1. *If $\sigma_n \to \sigma$ and $\sum u_n \epsilon T^*$, then $s_n \to \sigma$.*

THEOREM 9.2. *If $0 < \Delta < \pi/2$, $\epsilon > 0$, $0 < \alpha < 1 < \beta < \infty$, $k_0 > 0$, and $\psi_1, \psi_2, \cdots$ are such that when $k > k_0$,*

$$(9.21) \qquad \operatorname{minimum}_{\alpha k < p < q < \beta k} \mathcal{R}(s_q - s_p)e^{i(\psi_k + \delta_k)} \geqq -\epsilon \qquad |\delta_k| \leqq \Delta,$$

then for each complex number A,

$$(9.22) \qquad \limsup_{n \to \infty} |s_n - A| \leqq [\epsilon + C \limsup_{n \to \infty} |\sigma_n - A|] \operatorname{cosec} \Delta$$

where C is the greater of the numbers $(1 + \alpha)/(1 - \alpha)$ and $(\beta + 1)/(\beta - 1)$.

The first theorem is obviously a corollary of the second, for if $\sum u_n \epsilon T^*$ and $\sigma_n \to \sigma$, we can set $A = \sigma$ and choose ϵ as near 0 as we please. We give a direct proof of Theorem 9.2 which is similar to (and not essentially more difficult than) standard proofs of Tauberian theorems for classes smaller than T^*. Let $k' = [\alpha k]$ and $k'' = [\beta k]$. Then for each sufficiently great k the identity

$$(9.23) \qquad s_1 + s_2 + \cdots + s_k = (s_1 + s_2 + \cdots + s_{k'}) + (s_{k'+1} + \cdots + s_k)$$

can be written in the form

$$k\sigma_k = k'\sigma_{k'} + (k - k')s_k - \sum_{p=k'+1}^{k} (s_k - s_p),$$

and division by $k - k'$ gives

$$s_k = U_k' + V_k' \tag{9.25}$$

where

$$U_k' = \frac{k\sigma_k - k'\sigma_k'}{k - k'}; \qquad V_k' = \frac{1}{k - k'} \sum_{p=k'+1}^{k} (s_k - s_p). \tag{9.26}$$

If we set $L = \limsup |\sigma_n|$, then

$$\limsup | U_k' | \leqq \limsup [(k | \sigma_k | + k' | \sigma_{k'} |)/(k - k')]$$
$$\leqq L \lim [(k + k')(k - k')] = L(1 + \alpha)/(1 - \alpha) \leqq CL.$$

Using (9.21) gives for $k > k_0$ and $| \delta_k | \leqq \Delta$

$$\mathcal{R} V_k' e^{i(\psi_k + \delta_k)} \geqq \frac{1}{k - k'} \sum_{p=k'+1}^{k} (-\epsilon) = -\epsilon.$$

Hence

$$\liminf_{k\to\infty} \mathcal{R} s_k e^{i(\psi_k + \delta_k)} \geqq -\epsilon - CL. \tag{9.27}$$

Starting with the identity obtained by replacing k by k'' and k' by k in (9.23), a similar argument shows that

$$\limsup_{k\to\infty} \mathcal{R} s_k e^{i(\psi_k + \delta_k)} \leqq \epsilon + CL.$$

Hence

$$\limsup_{k\to\infty} | \mathcal{R} s_k e^{i(\psi_k + \delta_k)} | \leqq \epsilon + CL.$$

If we set $\varphi_k = \arg s_k$ so that $s_k = | s_k | \exp i\varphi_k$, this becomes

$$\limsup_{k\to\infty} | s_k | \, | \cos (\varphi_k + \psi_k + \delta_k) | \leqq \epsilon + CL.$$

As δ_k ranges over the interval $-\Delta \leqq \delta_k \leqq \Delta$, the angle $(\varphi_k + \psi_k + \delta_k)$ must assume a value which differs from each odd multiple of $\pi/2$ by at least Δ; hence

$$\limsup_{k\to\infty} | s_k | \cos (\pi/2 - \Delta) \leqq \epsilon + CL$$

and

$$\limsup | s_n | \leqq [\epsilon + \mathrm{C} \limsup | \sigma_n |] \operatorname{cosec} \Delta. \tag{9.28}$$

If s_n is replaced by $s_n - A$, then (9.21) still holds and σ_n is replaced by $\sigma_n - A$; therefore (9.28) holds when s_n and σ_n are replaced by $s_n - A$ and $\sigma_n - A$ respectively, and Theorem 9.2 is proved.

10. The Tauberian class of Pitt. In §9 we gave what is perhaps the simplest "direct" proof of C_1 Tauberian theorems for the class T^*. A more indirect proof of C_1 Tauberian theorems for the class T^* leads naturally to the general Tauberian classes of Pitt[14] and to generalizations of them.

Let $\sum u_n \in T^*$ and let $\epsilon > 0$. Then there exist $0 < \alpha < 1 < \beta$, a sequence $\psi_1, \psi_2, \cdots$, and an index k_0 such that when $k > k_0$ and $|\delta_k| \leqq \Delta$

$$\mathcal{R}(s_q - s_p)e^{i(\psi_k+\delta_k)} > -\epsilon \qquad \alpha k < p < q < \beta k.$$

In particular, when $k > k_0$ and $|\delta_k| \leqq \Delta$,

$$\mathcal{R}(s_p - s_k)e^{i(\psi_k+\delta_k)} > -\epsilon \qquad k < p < \beta k \tag{10.01}$$

and

$$\mathcal{R}(s_k - s_p)e^{i(\psi_k+\delta_k)} > -\epsilon \qquad \alpha k < p < k. \tag{10.02}$$

Our next step is to let A denote an arbitrary complex number; to replace s_p by $s_p - A$ and s_k by $s_k - A$ in (10.01) and (10.02); and then to set $\varphi_k = \arg(s_k - A)$ (our notation does not take into account the fact that φ_k depends on A) so that (10.01) and (10.02) become respectively

$$(s_p - A)e^{i(\psi_k+\delta_k)} > -\epsilon + |s_k - A| \cos(\varphi_k + \psi_k + \delta_k) \qquad k < p < \beta k$$

and

$$(s_p - A)e^{i(\psi_k+\delta_k+\pi)} > -\epsilon - |s_k - A| \cos(\varphi_k + \psi_k + \delta_k) \qquad \alpha k < p < k.$$

As δ_k ranges over the interval $-\Delta \leqq \delta_k \leqq \Delta$, the angle $(\varphi_k + \psi_k + \delta_k)$ must assume a value ω_k differing in absolute value from each odd multiple of $\pi/2$ by at least Δ; if possible, choose δ_k such that $|\delta_k| \leqq \Delta$ and ω_k lies in the first or fourth quadrants, and then set $\theta_k = \psi_k + \delta_k$; otherwise choose δ_k such that $|\delta_k| \leqq \Delta$ and ω_k lies in the second or third quadrants, and then set $\theta_k = \psi_k + \delta_k + \pi$. These choices of δ_k and θ_k give

$$\mathcal{R}(s_p - A)e^{i\theta_k} > -\epsilon + |s_k - A| \cos(\pi/2 - \Delta) \tag{10.03}$$

for each p in at least one of the two ranges $k < p < \beta k$ and $\alpha k < p < k$.

Let, for each ρ in the interval $0 < \rho \leqq 1$, T'_ρ denote the class of series $\sum u_n$ whose partial sums s_n satisfy the following condition: corresponding to each $\epsilon > 0$, there exist numbers α and β such that $\alpha < 1 < \beta$, a sequence $\theta_1, \theta_2, \cdots$, and an index k_0 such that when $k > k_0$, the inequality

$$\mathcal{R}s_p e^{i\theta_k} > -\epsilon + \rho |s_k| \tag{10.04}$$

holds for each p in at least one of the two ranges $\alpha k < p < k$ and $k < p < \beta k$. The argument leading to (10.04) establishes the following theorem in which Δ is the constant involved in the definition of T^*.

[14] H. R. Pitt, General Tauberian Theorems, Proc. London Math. Soc. (2) vol. 44 (1938), pp. 243–288. We refer to this paper as G.T.T. To compare the classes of G.T.T. with ours, it is necessary to associate a sequence s_n with a step function $s(x)$ defined by $s(x) = s_1 + s_2 + \cdots + s_{[x]}$ and make an exponential change of variable.

THEOREM 10.1. *If a series* $\sum u_n$ *with partial sums* s_n *is in class* T^*, *then for each constant A the series* $(u_1 - A) + u_2 + \cdots$ *with partial sums* $s_n - A$ *is in the class* T'_ρ *for which* $\rho = \cos(\pi/2 - \Delta)$.

It is clear that if $\rho' < \rho$, then $T_{\rho'} \supset T_\rho$. Let T' denote the union of the classes T'_ρ for $0 < \rho \leqq 1$. Then $T^* \subset T'$. If we modify the definition involving (10.04) by allowing ρ to depend on ϵ, we obtain a class P of series which is identical[15] with the class of series satisfying condition T of G.T.T. pp. 244–245.

DEFINITION 10.2. *Let P denote the class of series* $\sum u_n$ *whose partial sums* s_n *satisfy the following condition: corresponding to each* $\epsilon > 0$ *there exist numbers* ρ, α, *and* β *such that* $\rho > 0$ *and* $\alpha < 1 < \beta$; *a real sequence* $\theta_1, \theta_2, \cdots$; *and an index* k_0 *such that when* $k > k_0$ *the inequality*

$$\mathcal{R} s_p e^{i\theta_k} > -\epsilon + \rho \mid s_k \mid \tag{10.21}$$

holds for each p in at least one of the two ranges $\alpha k < p < k$ *and* $k < p < \beta k$.

Obviously $T' \subset P$. Hence the relations $T \subset T^* \subset T'$ and $T'_\rho \subset T'$ imply that each Tauberian oscillation theorem of G.T.T. which applies to series in P applies a fortiori to series in the classes T, T^*, T'_ρ, and T'.

11. The classes T''_E and T''. We now define still more general Tauberian classes by removing the requirement that (10.21) hold for each p in a range having k at one of the extremes. However we maintain (by restricting the sequences α_n and β_n) the requirement that (10.21) shall hold for each p in a range which is neither too short nor too remote from k. It will be clear that $P \subset T'' \subset T''_E$ for each $E > 0$.

DEFINITION 11.1. *Corresponding to each* $E > 0$, *let* T''_E *denote the class of series* $\sum u_n$ *whose partial sums* s_n *satisfy the following condition: there exist a constant* $\rho > 0$; *sequences* α_k *and* β_k *such that* $0 < \alpha_k < \beta_k$, $k\alpha_k \to \infty$, *and*

$$\limsup_{k\to\infty} (\beta_k + \alpha_k)/(\beta_k - \alpha_k) \equiv F < \infty; \tag{11.11}$$

a real sequence θ_k; *and an index* k_0 *such that when* $k > k_0$

$$\mathcal{R} s_p e^{i\theta_k} > -E + \rho \mid s_k \mid \qquad k\alpha_k \leqq p \leqq k\beta_k. \tag{11.12}$$

Let T'' *denote the intersection of the classes* T''_E *for* $E > 0$.

THEOREM 11.2. *If* $\sum u_n \in T''_E$ *then*

$$\limsup_{k\to\infty} \mid s_k \mid \leqq [E + F \limsup_{n\to\infty} \mid \sigma_n \mid]\rho^{-1} \tag{11.21}$$

where s_k *and* σ_n *are respectively the partial sums and the* C_1 *transform of* $\sum u_n$, *and the constants F and* $\rho > 0$ *are any set for which* (11.11) *and* (11.12) *hold.*

[15] It is easy to see that, apart from the difference in scale of the independent variable, the definitions differ only in appearance.

Using the notation of the definition of T''_E, let $k' = [k\alpha_k]$ and $k'' = [k\beta_k]$. Then the identity

$$(11.22) \qquad (s_1 + \cdots + s_{k''}) - (s_1 + \cdots + s_{k'}) = s_{k'+1} + \cdots + s_{k''}$$

leads, for k sufficiently great, to

$$(11.23) \qquad \begin{aligned} \mathcal{R}(k''\sigma_{k''} - k'\sigma_{k'})e^{i\theta_k} &= \mathcal{R} \sum_{p=k'+1}^{k'} s_p e^{i\theta_k} \\ &> \sum_{p=k'+1}^{k''} (-E + \rho|s_k|) = (k'' - k')(-E + \rho|s_k|) \end{aligned}$$

so that

$$(11.24) \qquad \rho\,|\,s_k\,| \leqq E + |\,k''\sigma_{k''} - k'\sigma_{k'}\,|/(k'' - k')$$

and our result follows easily by use of the crude inequality $|\,k''\sigma_{k''} - k'\sigma_{k'}\,| \leqq k''\,|\,\sigma_{k''}\,| + k'\,|\,\sigma_{k'}\,|$.

This proof of Theorem 11.2 is so simple and straightforward as to be almost trivial. Indeed it may seem that the general class T''_E, at which we arrived after several successive generalizations of the original Tauberian class of series for which $nu_n \to 0$, is one designed especially to make possible a simple and transparent proof. It is a significant fact that direct proofs of C_1 Tauberian theorems for the class of series for which $nu_n \to 0$ are simple and straightforward; that such proofs for intermediate classes such as the class of series for which $n\,|\,u_n\,| < K$ or the classes T or T^* are more devious and complicated; and that finally such proofs for the larger classes T''_E are again simple and straightforward.

Several corollaries of Theorem 11.2 are easily obtained. In the first place, if $\sum u_n \in T''_E$ and σ_n is bounded, then s_n must be bounded. If it is true that not only $\sum u_n \in T''_E$ but also the series $(u_1 - A) + u_2 + u_3 + \cdots$ is in class T''_E for each A (and, by Theorem 10.1, in particular if $\sum u_n \in T^*$), then we can replace s_n and σ_n in (11.21) by $s_n - A$ and $\sigma_n - A$ respectively to obtain

$$(11.25) \qquad \limsup_{k\to\infty} |\,s_k - A\,| \leqq [E + F \limsup_{n\to\infty} |\,\sigma_n - A\,|]\rho^{-1}.$$

If $\sigma_n \to \sigma$ and (11.25) holds, then we can set $A = \sigma$ to obtain

$$(11.26) \qquad \limsup_{k\to\infty} |\,s_k - \sigma\,| \leqq E/\rho.$$

If it is also true that $\sum u_n \in T''$ and that E and ρ can be chosen such that E/ρ is arbitrarily near 0 (and in particular if $\sum u_n \in T'_\rho$ for some $\rho > 0$), then (11.26) implies that $\sum u_n$ converges to σ.

12. Conclusion. In conclusion, we use the preceding study of Tauberian conditions as a basis for indication that the general problem of obtaining Tauberian convergence and oscillation theorems may be divided naturally into

four categories. Order among these categories is not significant, since each category is necessary for a complete theory.

I. The first problem is that of discovery of significant classes S of series $\sum u_n$ of such a character that one may decide by inspection of the terms of $\sum u_n$ (and preferably with no knowledge whatever concerning properties of the sequence s_n of partial sums of $\sum u_n$) whether $\sum u_n \epsilon S$.

II. The second problem is that of devising criteria to assist in showing that $\sum u_n \epsilon S$.

III. The third problem is that of showing that S is a subclass of a general class G of series for which Tauberian oscillation theorems can be proved by straightforward methods. To provide for Tauberian convergence theorems, it is also desirable to prove that if $\sum u_n \epsilon S$, then for each complex A the series $(u_1 - A) + u_2 + u_3 + \cdots$ is in class G.

IV. The fourth problem is that of establishing Tauberian theorems for the classes G.

Problems of type IV are the ones which depend essentially upon the particular method of summability used. The attention of this paper has been centered almost exclusively on problems of the first three types.

CORNELL UNIVERSITY,
ITHACA, N. Y.

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

ON DIVERGENCE PROPERTIES OF THE LAGRANGE INTERPOLATION PARABOLAS

BY P. ERDÖS

(Received November 13, 1939)

Throughout the present paper, $-1 < x_1^{(n)} < x_2^{(n)} < \cdots < x_n^{(n)} < 1$ denote the roots of the n-th Tchebicheff polynomial $T_n(x)$, and unless otherwise stated it is understood that the fundamental points of the Lagrange interpolation are the $x_i^{(n)}$.[1] It is well known that[2] there exists a continuous function whose interpolation parabolas diverge everywhere in $(-1, +1)$. In the present paper we prove that for $x_0 = \cos \frac{p}{q} \pi_1$, $p \equiv q \equiv 1 \pmod 2$, $(p_1, q) = 1$ there exists a continuous function $f(x)$ such that $L_n f(x_0) \to \infty$.[3] Turán and I[4] proved that this does not hold for any other point. In this direction Marcinkievicz[5] proved that if the fundamental points are the roots of $U_n(x) = T'_{n+1}(x)$ then for every continuous function $f(x)$ and every point x_0 there exists a sequence of integers $n_1 < n_2 < \cdots$ such that $L_{n_i}(f(x_0)) \to f(x_0)$. We remark that in the case of the Fourier series it is well known that there always exists a subsequence of the partial sums converging to $f(x_0)$. This fact may be of interest because there is often an analogous behaviour of the Lagrange interpolation parabola and the Fourier series.

First we prove some lemmas.

LEMMA 1.

$$x_i^{(m)} - x_j^{(n)} > \frac{1}{m^3}, \quad \textit{for } m \geqq n.$$

PROOF. Write

$$x_i^{(m)} = \cos \vartheta_i^{(m)}, \qquad \vartheta_i = \frac{2i-1}{2m}\pi.$$

Then we have

$$|x_i^{(m)} - x_j^{(n)}| > |\vartheta_i^{(m)} - \vartheta_j^{(n)}| \sin \frac{\pi}{2n} > \frac{\pi}{4n}\frac{\pi}{2mn} > \frac{1}{m^3} \text{ q.e.d.}$$

[1] For the employed notations see P. Erdös and P. Turán, Annals of Math., Vol. 38 (1937), p. 142–155. If there is no danger of confusion we will omit the upper index n.

[2] G. Grünwald, Annals of Math., Vol. 37 (1936), p. 908–918.

[3] $L_n(f(x))$ denotes the Lagrange interpolation parabola of $f(x)$.

[4] This result was stated in the Annals of Math., Vol. 38 (1937), p. 155 but there was a misprint.

[5] Acta Litt ac Scient. Szeged, Tom. 8, p. 127–130.

LEMMA 2. *Put* $x_0 = \cos \frac{p}{q}\pi$, $p \equiv q \equiv 1 \pmod 2$; *then constants* c_1 *and* c_2 *exist such that*

$$\min_{i=1,2,\cdots n} |x_0 - x_i^{(n)}| > \frac{c_1}{n}, \qquad |T_n(x_0)| > c_2.$$

PROOF.

$$|T_n(x_0)| = \cos\left(\frac{np}{q}\pi\right) \geqq \cos\left(\frac{\pi}{2} - \frac{\pi}{2q}\right) > c_2.$$

Put $x_j^{(n)} < x_0 < x_{j+1}^{(n)}$; then

$$\min_{i=1,2,\cdots n} |x_0 - x_i^{(n)}| > \frac{\pi}{2nq} \min\left(\sin \frac{2j-1}{2n}\pi, \sin \frac{2j+1}{2n}\pi\right) > \frac{c_1}{n}.$$

LEMMA 3.

$$\sum{}' |l_k^{(n)}(x_0)| < (\log n)^{\frac{1}{2}}$$

where $\sum'$ *indicates that the summation is extended only over the* $x_k^{(n)}$ *satisfying* $|x_k^{(n)} - x_0| > \frac{1}{(\log n)^{\frac{1}{2}}}$.

PROOF.

$$|l_k^{(n)}(x_0)| = \left|\frac{T_n(x_0)}{T_n'(x_k)(x_0 - x_k)}\right| < \frac{(\log n)^{\frac{1}{2}}}{n}$$

since $|T_n(x_0)| \leqq 1$ and $T_n'(x_k) = \frac{n}{\sqrt{(1 - x_k^2)}} \geqq n$, which proves the Lemma. Without loss of generality we may assume that $x_0 > 0$. Let $x_j^{(n)} < x_0 < x_{j+1}^{(n)}$. Now we prove

LEMMA 4. *Suppose* $0 < x_k^{(n)} < x_j^{(n)}$ $\left(\text{i.e.}, \frac{n}{2} < k < j\right)$; *then*

$$|l_k^{(n)}(x_0)| > \frac{c_3}{j-k}.$$

PROOF. We have

$$|l_k^{(n)}(x_0)| = \left|\frac{T_n(x_0)}{T_n'(x_k)(x_0 - x_k)}\right| \geqq \frac{c_2\sqrt{(1 - x_k^2)}}{n(x_0 - x_k)} > \frac{c_4}{n(x_{j+1} - x_k)},$$

by Lemma 2. Now $x_{j+1} - x_k < (j + 1 - k)\frac{\pi}{n} < \frac{c_5(j-k)}{n}$, which proves the Lemma.

LEMMA 5.

$$\sum_{(2k-1,n)=1} |l_k^{(n)}(x_0)| > c_6 \frac{\log n}{\log \log n}.$$

PROOF. By Lemma 4 we have

$$\sum_{(2k-1,n)=1} |l_k^{(n)}(x_0)| > \sum'' |l_k^{(n)}(x_0)| > c_3 \sum'' \frac{1}{j-k}$$

where the two dashes indicate that the summation is extended only over those k for which $(2k - 1, n) = 1$ and $\frac{n}{2} < k < j$. It is clear[6] that there are at least $c_7 n$ of the $x_k^{(n)}$ between 0 and $x_j^{(n)}$, thus

$$\sum'' \frac{1}{j-k} > \sum''' \frac{1}{j-k}$$

where the three dashes indicate that the summation is extended only over those k which satisfy $(2k - 1, n) = 1$ and $j - k < c_7 n$.

Denote by $\nu(n)$ the number of different odd prime factors of n. It is well known that $\nu(n) < c_8 \frac{\log n}{\log \log n}$. (This result is an immediate consequence of the prime number theorem, but can also be obtained in an elementary way.) The number of integers k satisfying $j - x < k < j$, $(2k - 1, n) = 1$ equals by the sieve of Eratosthenes

$$x - \sum_{p|n} \left[\frac{x}{p}\right]' + \sum_{pq|n} \left[\frac{x}{pq}\right]' - \cdots{}^{7} \geqq x \prod_{p|n} \left(1 - \frac{1}{p}\right) - 2^{\nu(n)}$$

$$> c_9 \frac{x}{\log \log n} - 2^{c_8 \log n / \log \log n} > c_{10} \frac{x}{\log \log n} \text{ for } x > \sqrt{n}, \quad (p \text{ odd})$$

since it is well known that $\prod_{p|n} \left(1 - \frac{1}{p}\right) > \frac{c_{11}}{\log \log n}$.[8] Thus

$$\sum''' \frac{1}{j-k} > \frac{c_{10}}{\log \log n} \sum_{c_7 n > r > \sqrt{n}} \frac{1}{r} > c_6 \frac{\log n}{\log \log n} \text{ q.e.d.}$$

THEOREM 1. *There exists a continuous function $f(x)$ such that $L_n(f(x_0)) \to \infty$.*

PROOF. Write

$$f(x) = \sum_{n=n_0}^{\infty} \frac{f_n(x)}{\sqrt{\log n}}.$$

---

[6] i.e. $|x_{r+1}^{(n)} - x_r^{(n)}| \leqq \frac{\pi}{n}$, $r = 1, 2, \cdots$.

[7] $\left[\frac{x}{p}\right]'$ denotes the number of the k's in the interval $j - x < k < j$ for which $2k - 1$ is divisible by p. It is clear that $\left[\frac{x}{p}\right]'$ differs from $\frac{x}{p}$ by less than 1.

[8] E. Landau, *Über den Verlauf der zahlentheoretischen Function.* Archiv der Math. und Phys., Ser. 3, Vol. 5, (1903), p. 86–91.

$f_n(x)$ is defined as follows:

$$f_n(x_k^{(n)}) = \text{signum } l_k^{(n)}(x_0) \quad \text{for} \quad (2k-1, n) = 1,$$

$$f_n\left(x_k^{(n)} \pm \frac{1}{2^{2^n}}\right) = 0,$$

in the intervals $\left(x_k^{(n)}, x_k^{(n)} + \frac{1}{2^{2^n}}\right)$ and $\left(x_k^{(n)}, x_k^{(n)} - \frac{1}{2^{2^n}}\right)$, $f_n(x)$ is linear and elsewhere $f_n(x) = 0$.

First we show that $f(x)$ is continuous. It suffices to show that

$$\sum_{n=n_0}^{\infty} \frac{f_n(x)}{\sqrt{\log n}}$$

is uniformly convergent, i.e. that

$$\sum_{n>n(\epsilon)} \frac{f_n(x)}{\sqrt{\log n}} < \epsilon.$$

If for a certain y, $f_n(y)$ and $f_m(y)$, $m > n$ are both different from 0, we have for a certain k_1 and k_2

$$|x_{k_1}^{(n)} - y| < \frac{1}{2^{2^n}}, \qquad |x_{k_2}^{(m)} - y| < \frac{1}{2^{2^m}},$$

i.e.

$$|x_{k_1}^{(n)} - x_{k_2}^{(m)}| < \frac{2}{2^{2^n}}.$$

But by Lemma 1

$$|x_{k_1}^{(n)} - x_{k_2}^{(m)}| > \frac{1}{m^3}$$

hence $2m^3 > 2^{2^n}$, i.e. $m > n^2$ for $n > 3$. Thus

$$\sum_{n>n(\epsilon)} \frac{f_n(x)}{\sqrt{\log n}} < \sum_{r>r(\epsilon)} \frac{1}{\sqrt{\log 2^{2^r}}} < \epsilon.$$

Put

$$\varphi_1(x) = \sum_{r=n_0}^{n-1} \frac{f_r(x)}{\sqrt{\log r}}, \qquad \varphi_2(x) = \sum_{r>n} \frac{f_r(x)}{\sqrt{\log r}}.$$

Then

$$L_n(f(x_0)) = L_n(\varphi_1(x)) + L_n\left(\frac{f_n(x)}{\sqrt{\log n}}\right) + L_n(\varphi_2(x)).$$

First we show that $L_n(\varphi_2(x)) = 0$. It will evidently suffice to show that for every k, $\varphi_2(x_k^{(n)}) = 0$ or that for $r > n$, $f_r(x_k^{(n)}) = 0$. If for a certain $r > n, f_r(x_k^{(n)}) \neq 0$ we have for a certain l

$$|x_k^{(n)} - x_l^{(r)}| < \frac{1}{2^{2^r}},$$

which does not hold for by Lemma 1 for $2^{2^r} > r^3$.

Next we estimate $L_n(\varphi_1(x))$. If for a certain $x_k^{(n)}$, $f_r(x_k^{(n)}) \neq 0$ then for a certain l

$$|x_k^{(n)} - x_l^{(r)}| < \frac{1}{2^{2^r}}$$

which by Lemma 1 means that

$$2^{2^r} < n^3 \qquad \text{or} \qquad r < 2 \log \log n \text{ for } n > n_0 .$$

Thus if for a certain $x_k^{(n)}$, $\varphi_1(x_k^{(n)}) \neq 0$ then by Lemma 2

$$|x_k^{(n)} - x_0| > \min_{i=1,2,\cdots r} |x_i^{(r)} - x_0| - \frac{1}{2^{2^r}} > \frac{c_1}{r} - \frac{1}{2^{2^r}} > \frac{1}{(\log n)^{\frac{1}{2}}} \quad \text{for} \quad r > n_0.$$

Thus by Lemma 3

$$L_n(\varphi_1(x_0)) < c_{12} \sum_{|x_k - x_0| > (\log n)^{-\frac{1}{2}}} |l_k^{(n)}(x_0)| < c_{12}(\log n)^{\frac{1}{2}}$$

Now by Lemma 5

$$L_n(f_n(x_0)) = \sum_{(2k-1,n)=1} |l_k^{(n)}(x_0)| > c_6 \frac{\log n}{\log \log n}$$

since for $(2k - 1, n) \neq 1$ $f_n(x_0) = 0$. Thus finally

$$L_n(f(x_0)) > c_6 \frac{(\log n)^{\frac{1}{2}}}{\log \log n} - c_{12} (\log n)^{\frac{1}{2}} \rightarrow \infty .$$

Similarly we could prove that a continuous $f(x)$ exists such that $L_n(f(x_0))$ converges to any given value.

THEOREM 2. *If $x_0 \neq \cos \frac{p}{q} \pi$, $p \equiv q \equiv 1 \pmod 2$ then there exists for every continuous $f(x)$ a sequence of integers $n_1 < n_2 < \cdots$ such that $L_{n_i}(f(x_0)) \rightarrow f(x_0)$.*

PROOF. First we prove that there exists a sequence if integers $n_1 < n_2 < \cdots$ such that $| T_{n_k}(x_0) | < \frac{c_{13}}{n}$. We need the following

LEMMA 6. *If $x_0 \neq \frac{p}{q}$, $p \equiv q \equiv 1 \pmod 2$, then the inequality*

$$\left| x_0 - \frac{2r - 1}{2n_k} \right| < \frac{c_{14}}{n_k^2}$$

has an infinite number of solutions.

PROOF. If x_0 is rational it is of the form $\frac{2r-1}{2n_k}$, thus the Lemma is trivial. Hence we may suppose that x_0 is irrational. It is well known that the equation $\left|x_0 - \frac{a}{b}\right| < \frac{1}{b^2}$ has an infinite number of solutions. If infinitely many of the b's are even the Lemma is proved, if not consider the least positive solution of

$$2a\,d - bf = 1.$$

Obviously $f \equiv 1 \pmod 2$ and $d < b$ thus

$$\left|x_0 - \frac{f}{2d}\right| \leqq \frac{1}{b^2} + \frac{1}{2bd} < \frac{c_{14}}{d^2}$$

which proves the Lemma.

If $\left|x_0 - \frac{2r-1}{2n_k}\right| < \frac{c_{14}}{n_k^2}$ we have

$$T_{n_k}(x_0) < \cos\left(\frac{\pi}{2} - \frac{c_{14}}{n_k}\right) < \frac{c_{13}}{n}.$$

Consider now a sequence of integers $n_1 < n_2 < \cdots$ with $\left|x_0 - \frac{2r-1}{n_k}\right| < \frac{c_{14}}{n_k^2}$. We are going to prove that $L_{n_k}(f(x_0)) \to f(x_0)$.

For $k \neq r$ we have

$$|\,l_k(x_0)\,| = \left|\frac{T_{n_k}(x_0)}{T'_{n_k}(x_k)(x_k - x_0)}\right| < \left|\frac{c_{13}}{n^2(x_k - x_0)}\right|.$$

Thus

$$\sum_{k\neq r} |\,l_k(x_0)\,| < \frac{c_{13}}{n^2}\sum_{k\neq r}\frac{1}{|\,x_k - x_0\,|} = o(1),^{9}$$

hence from

$$\sum_{k=1}^{n} l_k(x) \equiv 1$$

[9] We have

$$\sum_{k\neq r}\frac{1}{x_k - x_0} = \sum_{|x_k - x_0|\leqq(\log n)^{-1}}{}' \frac{1}{|\,x_k - x_0\,|}$$

$$+ \sum_{|x_k - x_0| > (\log n)^{-1}}{}' \frac{1}{|\,x_k - x_0\,|} < n\log n + cn\log n = o(n^2).$$

(The dash indicates that $k = r$ is omitted.)

it follows that

$$l_r(x_0) = 1 - o(1).$$

Thus

$$L_{n_k}(f(x_0)) = f(x_r)l_r(x_0) + \sum_{k \neq r} f(x_k)l_k(x_0) = (f(x_0) + \epsilon)[1 - o(1)] + o(1) \to f(x_0),$$

which proves Theorem 2.

On the other hand we can prove that for every x in $(-1, +1)$ there exists a continuous $f(x)$ such that

$$\lim_{n\to\infty} \frac{\sum_{m \leq n} L_m(f(x_0))}{n} = \infty \cdot$$

The proof is very similar to that of Theorem 1.

PRINCETON, N. J.

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

ADDITIVE SET FUNCTIONS AND VECTOR LATTICES

BY S. BOCHNER AND R. S. PHILLIPS

(Received March 29, 1940)

This paper arose out of an attempt to extend known results in the theory of countably additive set functions to the finitely additive case. The basic tool used in this investigation has been the vector lattice. We have found this approach helpful to the understanding of both additive set functions and vector lattices.

An important link between point functions and additive set functions is the theorem of Lebesgue that every set function which is absolutely continuous is the integral of a point function. Radon and Nikodym extended the theorem from ordinary Lebesgue measure to countably additive measure on general sets. If the measure is bounded this theorem states that in the Banach space of all absolutely continuous functions, step-functions are dense in the norm. It was shown recently by one of the authors (2) that the latter theorem remains valid for finitely additive Jordan measure in general; however in the course of the proof the more general case was reduced to the previous theorem of Lebesgue-Nikodym. In section II of the present note a new proof will be given. It will not presuppose any essential facts from the Lebesgue theory proper, being based on simple but important facts from the theory of vector lattices. These prerequisites are assembled in section I.

In section III we discuss some analogies between vector lattices and set functions; here again the approach rather than the result is new. The positive elements are shown to be additive set functions on the Boolean algebra of normal subspaces. As a consequence finitely additive set functions become completely additive on this algebra. It is significant to observe that this extension of the original algebra can also be expressed in terms of the stochastic distance. We further show that a vector lattice may be embedded in a second vector lattice having the same Boolean algebra of normal subspaces and possessing a unit element. Finally we consider briefly the question of a generalized base.

I. PROJECTIONS IN VECTOR LATTICES

A space L will satisfy the following five postulates:

I: L is a linear space with real scalars, and a relation $x > 0$ is defined on L.

II: If $x > 0$ and $y > 0$, then $x + y > 0$.

III: If $x > 0$ and α is a scalar, then $\alpha > 0$ implies $\alpha x > 0$ and conversely.

IV: Relative to the given order relation ($x > y$ means $x - y > 0$), L is a lattice.

We will write: $x \vee y$ for sup (x, y), $x \wedge y$ for inf (x, y); x^+ for $x \vee 0$, x^- for $-x \vee 0$, and $|x|$ for $x^+ + x^-$; the least upper bound (if it exists) of a set E will be denoted by $\bigvee_E x$ and the greatest lower bound by $\bigwedge_E x$.

V: If a set E is bounded above (below) then $\mathsf{V}_E x (\mathsf{\Lambda}_E x)$ exists.

Proof of the following theorems can be found in a paper by Freudenthal (3): $x = x^+ - x^-$; $|x| = x \vee -x = x^+ \vee x^-$; $x \to a + x$ and $x \to \alpha x$ ($\alpha > 0$) are lattice automorphisms; $x \to -x$ is a lattice anti-automorphism; $x \vee y + x \wedge y = x + y$; $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$; $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$; $\alpha > 0$ implies $\alpha \mathsf{V}_E x = \mathsf{V}_E \alpha x$, $\mathsf{V}_{E,F} x \vee y = \mathsf{V}_E x \vee \mathsf{V}_F y$, $\mathsf{V}_{E,F} x \wedge y = \mathsf{V}_E x \wedge \mathsf{V}_F y$, $\mathsf{V}_{E,F}(x + y) = \mathsf{V}_E x + \mathsf{V}_F y$ (if the least upper bounds on the right side exist); and dual statements for the greatest lower bounds.

x will be said to be *disjoint* from y if $|x| \wedge |y| = 0$. Given a subset Y of L, Y' will be the set of all $x \in L$ disjoint from every $y \in Y$. Iterating the process, we define Y'' to be $(Y')'$ and similarly for Y''' etc.

As Y'' contains all elements of L disjoint from every element of Y', $Y'' \supset Y$. If $Y \supset Z$, then $Y' \subset Z'$. In particular $Y''' \subset Y'$. It follows that $Y' = Y''' = Y^{v} = \cdots$ and $Y'' = Y^{iv} = Y^{vi} = \cdots$. The sets Y and Y' have at most the null vector in common.

A subset E of L which contains along with x an element v such that $2|x| \leqq v$ will be called a *D-set*.

A subspace of L satisfying postulates I, II, III, IV, and V for sets bounded above (below) in L is called *normal*[1] if it contains with x all y such that $0 \leqq |y| \leqq |x|$. By a *direct decomposition* of L is meant a choice of disjoint complementary normal subspaces—that is, of normal subspaces S and T such that $S \wedge T = 0$ (null vector) and L is the direct sum of S and T.

If $x \geqq 0$ and if Y is a D-set, then we define

$$x_Y = \mathsf{V}_{y \in Y} x \wedge y$$

to be the *projection* of x on Y. This is a fundamental notion for this paper. If Y is a normal subspace, then x_Y belongs to Y as it is the least upper bound of a set of elements in Y. If $x \geqq 0$ belongs to Y, then $x_Y = \mathsf{V}_{y \in Y} x \wedge y = x$. Further if $x \geqq 0$, then $x - x_Y \geqq 0$.

The purpose of this section will be to show that Y' and Y'' define a direct decomposition of L and to study the case in which Y contains but a single element.

THEOREM 1. *If Y is any subset of L, then Y' is a normal subspace.*

Suppose y is a fixed element of Y and the x_i belong to Y'. Y' is linear because $|x| \wedge |y| = 0$ implies $|\alpha x| \wedge |y| = 0$ and $0 \leqq |x_1 + x_2| \wedge |y| \leqq (|2x_1| \vee |2x_2|) \wedge |y| = (|2x_1| \wedge |y|) \vee (|2x_2| \wedge |y|) = 0$. As $Y' \subset L$, II and III are valid in Y'. To establish IV and V let E be a subset of Y' bounded above by $z \in L$. Then by a repeated application of $(u \vee v + u \wedge v = u + v)$, we obtain

$$\begin{aligned} 0 \leqq |\mathsf{V}_E x| \wedge |y| \leqq (\mathsf{V}_E x^+) \wedge |y| &= \mathsf{V}_E x^+ + |y| - \mathsf{V}_E x^+ \vee |y| \\ &= \mathsf{V}_E (x^+ \vee |y| + x^+ \wedge |y|) - \mathsf{V}_E x^+ \vee |y| \\ &\leqq \mathsf{V}_E (x^+ \vee |y|) + \mathsf{V}_E (x^+ \wedge |y|) - \mathsf{V}_E (x^+ \vee |y|) = 0. \end{aligned}$$

[1] Garrett Birkhoff (1) calls such a subspace a complete normal subspace; F. Riesz (6) denotes it by famille compléte.

The dual statement for boundedness from below can be similarly shown. Finally if $|w| \leqq |x|$, then $0 \leqq |w| \wedge |y| \leqq |x| \wedge |y| = 0$.

LEMMA 1. *If Y is a D-set and if $x \geqq 0 \in L$, then $x - x_Y \in Y'$.*

Let $y \geqq 0$ belong to Y. Then

$$\begin{aligned} 0 \leqq (x - x_Y) \wedge y &\leqq (x - x_Y) \wedge x \wedge y \\ &\leqq (x - x_Y) \wedge x_Y \\ &= -x_Y + x \wedge 2x_Y = -x_Y + x \wedge 2\mathsf{V}_{y \in Y}(x \wedge y) \\ &= -x_Y + \mathsf{V}_{y \in Y}(x \wedge 2x \wedge y) \\ &= -x_Y + x_Y = 0. \end{aligned}$$

As an obvious consequence of the lemma we have

$$(x - x_Y)_Y = 0$$

and this justifies the name "projection."

THEOREM 2. *If Y is a normal subspace of L, then $Y = Y''$.*

We can restrict ourselves to positive elements since if $x \in Y$, then $x^+ \in Y$ and $x^- \in Y$. If $x \geqq 0 \in Y''$, then by the lemma the element $x - x_Y$ belongs to Y'. However, it also belongs to Y'' since $0 \leqq x - x_Y \leqq x$ and Y'' is normal. Our element must therefore be the null element, and hence $x = x_Y$. But x was an arbitrary element of Y'', and x_Y belongs to Y.

THEOREM 3. *If Y is a normal subspace of L, then Y and Y' define a direct decomposition of L, namely*

$$x = (x_Y^+ - x_Y^-) + (x_{Y'}^+ - x_{Y'}^-).$$

Any decomposition $x = x_1 + x_2$ where $x_1 \in Y$ and $x_2 \in Y'$ is certainly unique because Y and Y' are linear and have only the null vector in common. It is therefore sufficient to show that for $x \geqq 0$, $x = x_Y + x_{Y'}$. By lemma 1, $x - x_Y \in Y'$. Hence $x_{Y'} \geqq (x - x_Y)_{Y'} = x - x_Y$. Since $Y = Y''$, we likewise have $x_Y \geqq x - x_{Y'}$. Therefore $x = x_Y + x_{Y'}$.

Frédéric Riesz (6) has demonstrated theorems 1, 2, 3 for the special case where L is the space of linear functionals on a semi-vector lattice.

If Y consists of a single element y, then Y'' will be called a *principal normal subspace* and be designated by $P(y)$.

From the definition, $P(y) = (y)''$, it follows that $x \in P(y)$ if and only if $|z| \wedge |y| = 0$ implies $|z| \wedge |x| = 0$. Now $x \in P(y)$ and $|x| \wedge |y| = 0$ implies $|x| = |x| \wedge |x| = 0$. Hence if $x \in P(y)$ and if $x \neq 0$, then $|x| \wedge |y| > 0$. In Freudenthal's terminology (3), $|y|$ is a unit element of $P(y)$.

THEOREM 4. *A necessary and sufficient condition that x belong to $P(y)$ is that $x = \mathsf{V}_n x^+ \wedge n|y| - \mathsf{V}_n x^- \wedge n|y|$.*

If x is of this form, x clearly belongs to $P(y)$. To prove the converse we need only consider $x \geqq 0 \in P(y)$. As the set $\{n|y|\}$ is a D-set, it follows from

lemma 1 that $x - \mathsf{V}_n x \wedge n \mid y \mid$ belongs to $P(y)'$. Being less than x, it also belongs to $P(y)$, and is therefore the null vector.

Theorem 4 shows that elements of $P(y)$ are approximable by elements bounded in $\mid y \mid$. The element y being fixed, boundedness in $\mid y \mid$ is the same as a Lipschitz condition in the case of point functions. If L is the space of additive set functions on a generalized Boolean algebra, $P(y)$ is the class of all functions absolutely continuous with respect to the set function y. Theorem 3 then gives a decomposition of the function into an absolutely continuous part and a singular part.

II. Absolutely Continuous Set Functions

We now consider a finitely additive Boolean algebra of subsets E of a given set G, and on it a (finitely additive) Jordan volume which will be denoted by $v(E)$ or $\mid E \mid$. We assume that $\mid G \mid = 1$.

We consider the Banach space V_1 of all set functions $F(E)$ of bounded variation relative to finite partitions $\delta = (E_\nu)$. We define $F > 0$ if $F(E) \geqq 0$ for all E and $F(E) > 0$ for some E. With this definition, V_1 has all properties I–V of a space L. The properties I–III are trivial. Let $[F_\alpha]$, α is an index, be a set of elements which are bounded from above by an element G. In order to show the existence of $\sup_\alpha F_\alpha$ we take for any set E any integer n, any set of indices $\alpha = \alpha_1, \cdots, \alpha_n$ (which need not be different), and any partition of E into disjoint sets $E_1, \cdots, E_n$, and we put

$$(1) \qquad F(E) = \sup_{n,\alpha_\nu,E_\nu} (F_{\alpha_1}(E_1) + \cdots + F_{\alpha_n}(E_n)).$$

Since $F_{\alpha_\nu}(E_\nu) \leqq G(E_\nu)$ and $G(E)$ is additive, we obtain $F(E) \leqq G(E)$ and therefore $F(E)$ is finite. It is not hard to see that $F(E)$ is additive. Also putting $\alpha_1 = \cdots = \alpha_n = \alpha$ we obtain $F(E) \geqq F_\alpha(E)$. Lastly if $H(E) \geqq F_\alpha(E)$ for all α, then $H(E) = H(E_1) + \cdots + H(E_n) \geqq F_{\alpha_1}(E_1) + \cdots + F_{\alpha_n}(E_n)$ and therefore $H(E) \geqq F(E)$. The inf. can be obtained in a dual fashion. Thus properties IV and V are also satisfied.

An element F of V_1 is called absolutely continuous if for any $\epsilon > 0$ there exists an $\eta > 0$ such that $\mid E \mid < \eta$ implies $\mid F(E) \mid < \epsilon$. The totality of such elements F will be denoted by AC. Obviously $v(E) \in AC$. An element G of V_1 is called singular if to any $\epsilon > 0$ there exists a set E_ϵ for which $\mid E_\epsilon \mid < \epsilon$ such that for any E, $\mid G(E - E_\epsilon) \mid < \epsilon$. The totality of all such elements G will be denoted by S. It is easily seen that AC and S are each linear spaces with properties I—IV.

Now, if $F \in AC$ and $G \in S$ and $F \geqq 0$ and $G \geqq 0$, then

$$\inf (F, G) = \inf_{E'+E''=E}(F(E') + G(E'')) = 0.$$

Therefore

$$S \wedge AC = 0 \qquad \text{(null function)}$$

and

$$S' \supset AC. \tag{2}$$

On the other hand if $G \geqq 0$, $G \in V_1$, then

$$\inf(G, v) = \inf_{E'+E''=E}(G(E') + v(E'')) = 0$$

if and only if $G \in S$. Therefore $S = (v)'$ and $(v)'' = S'$. Thus by (2) we obtain

$$(v)'' \supset AC. \tag{3}$$

Our next step is to show that

$$(v)'' \subset AC \tag{4}$$

and hence

$$(v)'' = AC = S'.$$

Now (4) asserts that any element $F \geqq 0$ for which

$$F = \lim_{n\to\infty} \inf(F, nv) \tag{5}$$

is absolutely continuous (theorem 4). Each element $F_n = \inf(F, nv)$ belongs to AC since $0 \leqq F_n(E) \leqq n \mid E \mid$. Furthermore $F_n(E)$ converges toward $F(E)$ uniformly in E since $0 \leqq F(E) - F_n(E) \leqq F(G) - F_n(G)$, and therefore $F(E)$ also belongs to AC. By theorem 1, AC and S are normal subspaces.

It is now very easy to prove the following theorem:

THEOREM 5. *Step functions are dense in AC.*

Since for $F \geqq 0$,

$$\| F - F_n \|_1 = F(G) - F_n(G)$$

our argument shows that "bounded" functions are dense in AC in the norm of V_1, and so it suffices to prove that step functions are dense in the manifold of bounded functions. Introducing the space V_2 (see (2)), since the norm of a bounded element is larger in V_2 than it is in V_1 it is sufficient to prove the

LEMMA 2. *Step functions are dense in V_2.*

If $f(x)$ is a bounded integrable point function and if for some E we put

$$a = \frac{1}{\mid E \mid} \int_E f(x)\, dv$$

then obviously

$$\int_E \mid f(x) - a \mid^2 dv = \int_E \mid f(x) \mid^2 dv - \mid a \mid^2 \cdot \mid E \mid.$$

We take a partition $\delta = (E_\nu)$, put $E = E_\nu$ and sum over ν. This will give

$$\| f - f_\delta \|^2 = \| f \|^2 - \| f_\delta \|^2. \tag{6}$$

Now let F be an arbitrary element of V_2, δ a fixed partition, and δ' any partition $> \delta$. Since $F_{\delta'}$ (see (2)) is a step function, and $F_\delta = (F_\delta)_{\delta'}$ we obtain from (6)

$$\| (F - F_\delta)_{\delta'} \|^2 = \| F_{\delta'} \|^2 - \| F_\delta \|^2.$$

Taking the limit with respect to δ' we obtain by the definition of the norm

$$\| F - F_\delta \|^2 = \| F \|^2 - \| F_\delta \|^2.$$

Thus

$$\lim_\delta \| F - F_\delta \| = 0$$

and therefore F is the limit in the norm of step functions F_δ.

A Banach space X will be said to possess a *generalized base* if there exists a directed set of linear transformations $U_\delta(x)$ on X to a finite dimensional subspace of X such that (1) $\lim_\delta U_\delta(x) = x$ for all $x \in X$, and (2) $\| U_\delta \| \leqq M$ for all δ.

If X is the space AC and $\delta = (E_\nu)$ is a finite partition such that $v(E_\nu) \neq 0$ for any ν, then define

$$U_\delta(F) = \sum\nolimits_\delta \frac{F(E_\nu)}{v(E_\nu)} \cdot v(E \cdot E_\nu).$$

It is clear from the proof of theorem 5 that $\lim_\delta U_\delta(F) = F$ in the norm topology. Further $\| U_\delta(F) \|_1 \leqq \| F \|_1$. Therefore AC possess a generalized base.

III. The Boolean Algebra of Normal Subspaces

The class **B** of all normal subspaces of a vector lattice L with properties I–V is a Boolean algebra. This known fact and further information concerning **B** will be derived in the present section from properties of projections.

In **B** the element L is the 1 element, the normal subspace consisting of only the null vector is the 0 element, and for a set $E = [A]$ of **B**, $\wedge_E A$ is the set intersection and $\vee_E A$ is the smallest element of **B** containing all elements A of E. Also if A is an arbitrary element of **B**, then the derived element A', as defined in section I, is the complement of A in **B** in the sense that $A \wedge A' = 0$, $A \vee A' = 1$, and $A'' = A$. It remains only to prove that **B** satisfies the distributive law. This will be done in theorem 8.

In the next three theorems, a fixed class of elements $[A_\alpha]$ of B will be considered, and as an abbreviation we will denote by P the normal subspace $\vee_\alpha A_\alpha$, and by R the set sum of all elements of the A_α. Obviously $P \supset R$, and therefore $P' \subset R'$ and $P'' = P \supset R''$. However since P is the smallest normal subspace containing R, $R'' \supset P$. Therefore $P = R''$ and $P' = R'$.

Theorem 6. *If* $[A_\alpha]$ *is any class of elements of* B, *then*

$$(\vee_\alpha A_\alpha)' = \wedge_\alpha A_\alpha' \tag{7}$$

and

$$(\wedge_\alpha A_\alpha)' = \vee_\alpha A_\alpha'. \tag{8}$$

The relation $P' = R'$ implies that the following statements are equivalent: $x \in (\vee_\alpha A_\alpha)'$; $|x| \wedge |a_\alpha| = 0$ for all elements of R; $x \in A_\alpha'$ for all α; and $x \in \wedge_\alpha A_\alpha'$. This proves (7). (8) follows from (7) by taking the complement on both sides and replacing A_α by A_α'.

THEOREM 7. *If* $[A_\alpha]$ *is any class of elements of* **B** *and if* $x \geqq 0$, *then*

$$x_{\vee A_\alpha} = \vee_\alpha x_{A_\alpha} \tag{9}$$

and

$$x_{\wedge A_\alpha} = \wedge_\alpha x_{A_\alpha}. \tag{10}$$

Now R is clearly a D-set. Therefore by lemma 1, $x - x_R \in R' = P'$. Hence $x_{P'} = x - x_P \geqq x - x_R$. But $P \supset R$ implies $x_P \geqq x_R$, so that $x_P = x_R$. We have

$$x_R = \vee_{\alpha, A_\alpha} x \wedge a_\alpha = \vee_\alpha[\vee_{A_\alpha} x \wedge a_\alpha] = \vee_\alpha x_{A_\alpha}.$$

This proves (9). (10) follows from (9) in the following way:

$$x_{\wedge A_\alpha} = x_{(\vee A_\alpha')'} = x - x_{\vee A_\alpha'} = x - \vee_\alpha x_{A_\alpha'} = x - \vee_\alpha(x - x_{A_\alpha})$$
$$= x - (x - \wedge_\alpha x_{A_\alpha}) = \wedge_\alpha x_{A_\alpha}.$$

THEOREM 8. **B** *is a complete Boolean algebra. If* $[A_\alpha]$ *and* **B** *are elements of* **B**, *then*

$$(\vee_\alpha A_\alpha) \wedge B = \vee_\alpha(A_\alpha \wedge B)$$

and

$$(\wedge_\alpha A_\alpha) \vee B = \wedge_\alpha(A_\alpha \vee B).$$

In fact for $x \geqq 0$, we have by theorem 7

$$x_{\vee(A_\alpha \wedge B)} = \vee_\alpha x_{A_\alpha \wedge B} = \vee_\alpha(x_{A_\alpha} \wedge x_B) = \vee_\alpha x_{A_\alpha} \wedge x_B = x_{\vee A_\alpha} \wedge x_B = x_{(\vee A_\alpha) \wedge B}.$$

Before giving the next theorem we need a *definition*. The *sum* of any set E of non-negative elements of L will be defined as follows: For a given ordering x_α of E, suppose $\sum_\lambda x_\alpha$ to be defined for all $\lambda < \mu$. Define $\sum_\lambda x_\alpha = \vee_{\lambda<\mu}[\sum_\lambda x_\alpha] + x_\mu$. Continuing this process until the set is exhausted, obtains an element $\sum x_\alpha \in L$. Suppose that x_α' is another ordering of E. We next show that $\sum x_\alpha' = \sum x_\alpha$. It is easily shown that for any finite set $\pi = (\alpha_1, \cdots, \alpha_n)$, $\sum_\pi x_\alpha' \leqq \sum x_\alpha$. Suppose now that for all $\lambda < \mu$, $\sum_\lambda x_\alpha' \leqq \sum x_\alpha - \sum_\pi x_\alpha'$ where π is any finite set of α's greater than or equal to μ. Hence $\vee_{\lambda<\mu}[\sum_\lambda x_\alpha'] \leqq \sum x_\alpha - \sum_\pi x_\alpha'$. As μ could also belong to π, this proves the induction. It follows that $\sum x_\alpha' \leqq \sum x_\alpha$. Similarly we could show that $\sum x_\alpha' \geqq \sum x_\alpha$. Our definition of sum therefore depends only on the set E. We will designate this sum by $\sum_E x$.

On the basis of the previous theorem 8 we now have the

THEOREM 9. *If $x \geqq 0$, then x_A defines a lattice homomorphism of* **B** *on L. In particular if $[A_\alpha]$ is an arbitrary class of disjoint elements of* **B**, *then*

$$x_{\vee A_\alpha} = \sum_\alpha x_{A_\alpha} .$$

THEOREM 10. *If L has a unit $y > 0$, that is $L = P(y)$, then the homomorphism which is defined by y_A (see theorem 9) is an isomorphism.*

Also, for each $A \in \mathbf{B}$, $A = P(y_A)$ and thus $e = y_A$ can be written in the form $e = y_{P(e)}$. An element e can be written in the form y_A if and only if

$$(11) \qquad 0 \leqq e \leqq y \quad \text{and} \quad e \wedge y - e = 0.$$

PROOF: By theorem 9, $A \to y_A$ is a homomorphism. Now if $A \to 0$, then $a \in A$, $a \geqq 0$ implies $y \wedge a = 0$. Since y is a unit this implies $a = 0$ and hence $A = 0$. Thus $A \to y_A$ is an isomorphism.

Since $y_A \in A$, we also have $P(y_A) \subset A$. Thus $P(y_A) = \varphi(A)$ is a function from B to B, and $\varphi(A)$ is always a (proper or improper) part of A. However, $\varphi(P(y_A)) = P(y_A)$, and if A_1, A_2 are disjoint then so are $\varphi(A_1)$ and $\varphi(A_2)$. Therefore $\varphi(A) = A$; that is $A = P(y_A)$. Finally (11) is fulfilled for $e = y_A$. If (11) holds then $e \in P(e) = A$ and $y - e \in A'$, so that $e = y_A$ by theorem 3.

As a corollary to theorem 10 we have

THEOREM 11. *If L has a unit $y > 0$, then the Boolean algebra B of normal subspaces A is isomorphic with the Boolean algebra E of elements e of L for which* (11) *holds.*

Consider a generalized finitely additive Boolean algebra T of elements τ, a finitely additive positive set function $y(\tau)$ on T, and the class L of all finitely additive set functions on T absolutely continuous with respect to $y(\tau)$. Our theory suggests a method of extending T modulo null sets to a sigma Boolean algebra on which $y(\tau)$ is completely additive. Now clearly y is a unit for $L = P(y)$. Each $\tau_0 \in T$ defines a function $y(\tau \cdot \tau_0)$ which in the notation of theorem 11 belongs to E. If $x > 0 \in L$, then it is easily shown that $x_{P(y(\tau \cdot \tau_0))} = x(\tau \cdot \tau_0)$. All sets of y-measure zero correspond to 0 in the Boolean algebra of normal subspaces. Finally it is necessary to add certain ideal elements of E which do not correspond to elements of T modulo null sets. By theorem 9 the set functions x_A are completely additive on B. In particular their values on the unit of T are completely additive. On sets which do correspond to τ_0, this value is precisely $x(\tau_0)$. It is clear that by treating x^+ and x^- in this way $x(\tau) \in L$ becomes a completely additive set function on the extended algebra E.

Kakutani has obtained a similar result, namely, the given Boolean algebra can be embedded in a sigma field of point sets such that the finitely additive function can be extended on this new field to be completely additive. The new field is in this case difficult to determine.

Now our extension modulo null sets of T to E is the smallest possible extension for which $y(\tau)$ becomes completely additive. For suppose T' were another such extension. In this case the elements of E' defined by (11) for the new lattice L' do correspond to elements of T'. As T' contains T modulo null sets, L' is lattice homomorphic with $L = P(y)$. Therefore $E' \supset E$.

We define the stochastic distance $\rho(\tau_1, \tau_2) = y(\tau_1 \cdot \tau_2^1) + y(\tau_1^1 \cdot \tau_2)$. If we complete T in the usual manner by taking Cauchy sequences, we obtain a sigma field which is an extension modulo null sets of T and on which y becomes completely additive. As the metric extension of a sigma field modulo null sets relative to a completely additive set function is precisely the original field modulo null sets, it follows that the metric extension is likewise the smallest possible extension modulo null sets of T. Therefore E is also the metric extension of T modulo null sets.

Now if L has no unit it is possible to obtain (by means of the choice axiom) a set of principal normal subspaces $A_\alpha = P(y_\alpha)$ such that $\vee_\alpha A_\alpha = L$, $A_\alpha \wedge A_\beta = 0$ if $\alpha \neq \beta$, and $y_\alpha > 0$. Putting $x_\alpha = x^+_{A_\alpha} - x^-_{A_\alpha}$ each element x of L can then be represented by the set of components $[x_\alpha]$, $x_\alpha \in A_\alpha$. Not all possible combinations of components $[x_\alpha]$ will occur in the representation.

We now define a new vector lattice L^* to consist of *all* possible elements $x = [x_\alpha]$, $x_\alpha \in A_\alpha$. Defining the lattice operations in the obvious way, the new lattice satisfies postulates I–V, possesses a unit element, namely $[y_\alpha]$, and the original lattice L can be embedded into it.

The Boolean algebra $\mathbf{B}^*$ of normal subspaces of L^* is lattice isomorphic with the Boolean algebra $\mathbf{B}$ of normal subspaces of L. $B^* \in \mathbf{B}^*$ is the direct product of the normal subspaces $B_\alpha \subset A_\alpha$. We correspond $B^* \in \mathbf{B}^*$ to $\vee_\alpha B_\alpha \in \mathbf{B}$ and $A \in \mathbf{B}$ to $[A \wedge A_\alpha] \in \mathbf{B}^*$. By theorem 8, $(\vee_\alpha B_\alpha) \wedge A_\alpha = B_\beta$ and $\vee_\alpha(A \wedge A_\alpha) = A$. Hence this is a one to one correspondence. The correspondence clearly preserves order.

We wish finally to consider the question of a generalized base (see section II) in vector lattices which possess a norm relative to which they are complete. We suppose given a set of normal subspaces A_α each possessing a generalized base U^α_δ where the $\| U^\alpha_\delta \|$ are uniformly bounded. We suppose further that $\vee_\alpha A_\alpha = L$, $A_\alpha \wedge A_\beta = 0$ if $\alpha \neq \beta$, and that the norm of x $\eta(x) = \sum_\alpha \eta(x_{A_\alpha})$. These conditions are satisfied by vector lattices of type (L) (1) among which is the space V_1. It now follows (5) that L itself possesses a generalized base.

PRINCETON UNIVERSITY
THE INSTITUTE FOR ADVANCED STUDY

REFERENCES

1. Garrett Birkhoff, *Dependent Probabilities and Spaces* (L), Proceedings of the National Academy of Sciences, vol. 24, pp. 154–159, 1938.
2. S. Bochner, *Additive Set Functions on Groups*, Annals of Mathematics, vol. 40, pp. 769–799, 1939.
3. Hans Freudenthal, *Teilweise Geordnete Moduln*, Proceedings Akademie van Wetenschappen Amsterdam, vol. 39, pp. 641–651, 1936.
4. L. Kantorovitch, *Lineare Halbgeordnete Raume*, Recueil Mathematique, vol. 2, pp. 121–168, 1937.
5. R. S. Phillips, *On the Space of Completely Additive Set Functions*, Duke Mathematical Journal.
6. Frédéric Riesz, *La théorie générale des opérations linéaires*, Annals of Mathematics, vol. 41, pp. 174–206, 1940.

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

FREE LATTICES[1]

BY PHILIP M. WHITMAN

(Received October 18, 1939)

1. Introduction

A **lattice** (sometimes called a structure) is a partially ordered set of elements each two of which have a greatest lower and a least upper bound, denoted $A \cap B$ and $A \cup B$, read "A meet B" and "A join B." Postulates[2] for partial ordering are that $\leqq$, defined between some pairs of elements, satisfy

$$A \leqq A \text{ for all } A; \tag{1}$$

$$\text{if } A \leqq B \text{ and } B \leqq A, \text{ then } A = B; \tag{2}$$

$$\text{if } A \leqq B \text{ and } B \leqq C, \text{ then } A \leqq C. \tag{3}$$

A lattice is said to be **generated** by a set of elements X_i (the "generators") if it consists of the X_i and their finite combinations by $\cap$ and $\cup$– e.g., $X_1 \cup X_2$, $\{[X_1 \cup X_2] \cap X_3\} \cup X_1$ –sometimes known as "lattice polynomials."

However, these polynomials do not all constitute distinct elements; for instance, $X_i \cap X_i = X_i$ in any lattice by definition of $\cap$; furthermore, in a specific lattice we may have additional rules of equality; e.g. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

The **free lattice** generated by the X_i is[3] a lattice generated by them in which there are no laws of equality except those derivable from the postulates for a lattice. This is the most general lattice generated by the X_i, in the sense that every other can be obtained from it by a homomorphism– determined by the additional rules of equality.

We are concerned here with the internal structure of a free lattice. Given two polynomials, how can we find which of $\leqq$, $=$, $\geqq$ (if any) hold between them in the free lattice (equivalently, in all lattices)? This question is answered in §2. In §3 we show that given a polynomial, there is a shortest polynomial

[1] Presented to the American Mathematical Society Sept. 5, 1939.

[2] Equivalently, we may postulate for a lattice the identities $L1$: $X \cup X = X \cap X = X$; $L2$: $X \cup Y = Y \cup X, X \cap Y = Y \cap X$; $L3$: $X \cup (Y \cup Z) = (X \cup Y) \cup Z$, $X \cap (Y \cap Z) = (X \cap Y) \cap Z$; $L4$: $X \cup (X \cap Y) = X \cap (X \cup Y) = X$, note the theorem that $X = X \cap Y$ if and only if $Y = X \cup Y$, and define $X \leqq Y$ as equivalent to $X = X \cap Y$ or $Y = X \cup Y$. Cf. Ore, *On the Foundation of Abstract Algebra I*, Annals of Math., 36 (1935), p. 409.

[3] Its existence is guaranteed by a theorem of universal algebra; cf. G. Birkhoff, *On the Structure of Abstract Algebras*, Proc. Camb. Phil. Soc., 31 (1935), pp. 440-1. Compare free groups.

equal to it in the free lattice (which can be found by a definite procedure), and in §4 it is shown that certain elements cover others.

The author is indebted to Prof. Garrett Birkhoff for many helpful suggestions.

2. Conditions for $A \leqq B$ in a free lattice

Given two elements A and B, we should like to know: is $A \leqq B$ in the free lattice? Sometimes this is obviously true; e.g. $X_1 \leqq X_1 \cup X_2$. Sometimes we can show it false, for by the abovementioned homomorphism, $A \leqq B$ in the free lattice implies $A \leqq B$ in every lattice with the same generators. Hence if we can exhibit a specific lattice where $A \nleqq B$, then $A \nleqq B$ in the free lattice. But this is a trial and error method; we should like some definite procedure for settling the question.

THEOREM 1. *In the free lattice generated by a set of elements* X_i,

(4) $X_i \leqq X_j$ if and only if $i = j$;

(5) *recursively,* $A \leqq B$ *if and only if one or more of the following hold:*

(a) $A \equiv A_1 \cup A_2$ *where* $A_1 \leqq B$ *and* $A_2 \leqq B$,

(b) $A \equiv A_1 \cap A_2$ *where* $A_1 \leqq B$ *or* $A_2 \leqq B$,

(c) $B \equiv B_1 \cup B_2$ *where* $A \leqq B_1$ *or* $A \leqq B_2$,

(d) $B \equiv B_1 \cap B_2$ *where* $A \leqq B_1$ *and* $A \leqq B_2$.

NOTE. In ($5a$) it is permissible that A_1 be itself a join; etc.

We see that this is the sort of condition desired, for, given A and B, we need look only at B and part of A, and A and part of B. Since the elements are the *finite* combinations of the generators, this process will eventually end with (4).

PROOF. These conditions are obviously sufficient; to prove them necessary we proceed by a series of definitions and lemmas.

DEFINITION OF $\subset$. **(6)** $X_i \subset X_j$ if and only if $i = j$; **(7)** $A \subset B$ if and only if one or more of ($5a$–d) are true with $\leqq$ replaced by $\subset$.

(8) DEFINITION. $A \supset B$ if and only if $B \subset A$.

NOTE. The set of definitions (6)–(8) is self-dual; i.e., if $\cap$ and $\cup$, $\subset$ and $\supset$ are interchanged the set remains the same. In particular, ($7a$) and ($7d$) are dual, ($7b$) and ($7c$) are dual, and (6) is self-dual. This property will enable us to omit many cases in proofs, where we need only make these same changes throughout.

(9) DEFINITION. $A \cong B$ if and only if $A \subset B$ and $A \supset B$.

(10) DEFINITION. The **length** of A, denoted $L(A)$, is the total number of X's appearing in A, counting repetitions; e.g., $L(X_1) = 1$, $L(X_1 \cup X_1) = 2$, $L(\{[X_1 \cup X_2] \cap X_3\} \cup X_1) = 4$.

With $\leqq$ as inclusion relation and $=$ as equality the combinations of the X_i form the free lattice; we now show that they also form a lattice with $\subset$ and $\cong$ in these roles; cf. (16).

(11) LEMMA. **(a)** $A \subset A$, **(b)** $A \subset A \cup B$, **(c)** $A \supset A \cap B$ *for all* A, B. *Thus* $A \cup B$ *is an upper bound to* A *and* B *under* $\subset$.

PROOF by induction on $L(A)$. ($11a$) is true for $L(A) = 1$, by (6). If ($11a$) is true for $L(A) \leqq m$, then so are ($11b, c$) by ($7c, b$). Hence ($11a$) holds for

$L(A) = m + 1$, for say $A \equiv A_1 \cup A_2$ (dually if $A \equiv A_1 \cap A_2$); then $A_1 \subset A$, $A_2 \subset A$ by induction; $\therefore A \equiv A_1 \cup A_2 \subset A$ by (7*a*).

(12) Lemma. **(a)** *$A_1 \cap A_2 \subset X_j$ if and only if $A_1 \subset X_j$ or $A_2 \subset X_j$.*
(b) *$X_j \subset B_1 \cup B_2$ if and only if $X_j \subset B_1$ or $X_j \subset B_2$.*
(c) *$A_1 \cup A_2 \subset B$ if and only if $A_1 \subset B$ and $A_2 \subset B$.*
(d) *$A \subset B_1 \cap B_2$ if and only if $A \subset B_1$ and $A \subset B_2$.*
(e) *$A_1 \cap A_2 \subset B_1 \cup B_2$ if and only if $A_1 \cap A_2 \subset B_1$ or $A_1 \cap A_2 \subset B_2$ or $A_1 \subset B_1 \cup B_2$ or $A_2 \subset B_1 \cup B_2$.*

Proof. "If" is obviously true by (7). "Only if" is proved by induction on $m = L(A) + L(B)$. It is true for $m \leqq 2$ vacuously. Assume (12) true for $m \leqq k - 1$; then for $m = k$,

(12*a*) $A_1 \cap A_2 \subset X_j$. Then $A_1 \subset X_j$ or $A_2 \subset X_j$, as desired, by (7*b*), the only part of (7) which applies.

(12*c*) $A_1 \cup A_2 \subset B$. *Case* 1. $A_1 \cup A_2 \subset X_j$. Then $A_1 \subset X_j$, $A_2 \subset X_j$ by (7*a*). *Case* 2. $A_1 \cup A_2 \subset B_1 \cup B_2$. Then by (7*a*, *c*), (*i*) $A_1 \subset B$ and $A_2 \subset B$ or (*ii*) $A_1 \cup A_2 \subset B_1$ or (*iii*) $A_1 \cup A_2 \subset B_2$. If (*ii*) $A_1 \cup A_2 \subset B_1$, then $A_1 \subset B_1$ and $A_2 \subset B_1$ by induction (12*c*); $\therefore A_1 \subset B_1 \cup B_2$, $A_2 \subset B_1 \cup B_2$ by (7*c*), as desired. Likewise if (*iii*) holds the lemma does, and if (*i*), then the proof is immediate. *Case* 3. $A_1 \cup A_2 \subset B_1 \cap B_2$. Then (*i*) $A_1 \subset B_1 \cap B_2$ and $A_2 \subset B_1 \cap B_2$ or (*ii*) $A_1 \cup A_2 \subset B_1$ and $A_1 \cup A_2 \subset B_2$, by (7*a*, *d*). If (*i*), Q.E.D. If (*ii*), then $A_1 \subset B_1$, $A_2 \subset B_1$, $A_1 \subset B_2$, $A_2 \subset B_2$ by induction (12*c*); $\therefore A_1 \subset B_1 \cap B_2$, $A_2 \subset B_1 \cap B_2$ by (7*d*).

(12*b*, *d*) dually.

(12*e*) $A_1 \cap A_2 \subset B_1 \cup B_2$. Then $A_1 \subset B_1 \cup B_2$ or $A_2 \subset B_1 \cup B_2$ or $A_1 \cap A_2 \subset B_1$ or $A_1 \cap A_2 \subset B_2$ by (7*b*, *c*). Q.E.D.

(13) Lemma. *If $A \subset B$ and $B \subset C$, then $A \subset C$.*

Proof by induction on $m = L(A) + L(B) + L(C)$. True for $m = 3$ by (6). Induction:

Case 1. *Non-meet* $\subset B \subset C$.

Case 1*a*. $X_i \subset X_j \subset C$. $\therefore i = j$ by (6). $\therefore X_i \equiv X_j$. $\therefore X_i \subset C$.

Case 1*b*. $X \subset B_1 \cup B_2 \subset C$. $\therefore X \subset$ some B_i by (12*b*). $B_i \subset C$ by (12*c*). $\therefore X \subset C$ by induction.

Case 1*c*. $X_i \subset B_1 \cap B_2 \subset X_j$. Part of dual of case 1*b*.

Case 1*d*. $X \subset B_1 \cap B_2 \subset C_1 \cup C_2$.

$\therefore B_1 \cap B_2 \subset$ some C_i	or some $B_i \subset C_1 \cup C_2$ by (12*e*).
$\therefore X \subset C_i$ by induction.	$X \subset B_i$ by (12*d*).
$\therefore X \subset C_1 \cup C_2$ by (12*b*) or (7*c*).	$\therefore X \subset C_1 \cup C_2$ by induction.

Case 1*e*. $A_1 \cup A_2 \subset B \subset C$. $\therefore A_i \subset B$ (all i) by (12*c*). $\therefore A_i \subset C$ (all i) by induction. $\therefore A_1 \cup A_2 \subset C$ by (12*c*).

Case 1*f*. $X \subset B_1 \cap B_2 \subset C_1 \cap C_2$. Part of dual of case 1*e*.

Case 2. *Meet* $\subset B \subset C$.

Case 2*a*. $A_1 \cap A_2 \subset B \subset$ non-join. Part of dual of case 1.

Case 2*b*. $A_1 \cap A_2 \subset X \subset C_1 \cup C_2$. $\therefore$ some $A_i \subset X$ by (12*a*). $\therefore A_i \subset C_1 \cup C_2$ by induction. $\therefore A_1 \cap A_2 \subset C_1 \cup C_2$ by (12*e*).
Case 2*c*. $A_1 \cap A_2 \subset B_1 \cup B_2 \subset C_1 \cup C_2$.

$\therefore A_1 \cap A_2 \subset$ some B_i	or some $A_i \subset B_1 \cup B_2$ by (12*e*).
$B_i \subset C_1 \cup C_2$ by (12*c*).	$\therefore A_i \subset C_1 \cup C_2$ by induction.
$\therefore A_1 \cap A_2 \subset C_1 \cup C_2$ by induction.	$\therefore A_1 \cap A_2 \subset C_1 \cup C_2$ by (12*e*).

Case 2*d*. $A_1 \cap A_2 \subset B_1 \cap B_2 \subset C_1 \cup C_2$. Dual of case 2*c*.

(14) Lemma. *$A \cup B$ is the least upper bound to A and B under $\subset$; dually for $A \cap B$.*

For it is *an* upper bound by (11), and if $A \subset C$ and $B \subset C$, then $A \cup B \subset C$ by (7*a*).

(15) Lemma. *$\cong$ is an equality relation.*[4]

(16) Lemma. *The finite combinations of the X_i by $\cap$ and $\cup$ form a lattice, generated by the X_i, with $\subset$ as inclusion relation and $\cong$ as equality and $A \cup B$, $A \cap B$ as least upper, greatest lower bounds to A and B.*

Proof. (1), (2), (3) are satisfied [(11), (9), (13)], and $A \cup B$ is least upper bound by (14).

Proof of theorem 1. Any other lattice generated by the X_i is a homomorphic image of the free lattice, hence $\leqq$ in a free lattice is sufficient for $\subset$. But by definition of $\cup$ as least upper bound and induction, $\subset$ is sufficient for $\leqq$ and hence $\cong$ for $=$. Therefore $\subset$ and $\leqq$ are equivalent in the free lattice. Theorem 1 then follows from (6) and (7).

Now denote $A_1 \cup A_2 \cup \cdots \cup A_n$ by $\sum_1^n A_i$, $A_1 \cap \cdots \cap A_n$ by $\prod_1^n A_i$, or simply $\sum A_i$, $\prod A_i$ if there is no confusion.

(17) Corollary. **(a)** *$\prod A_i \subset X_j$ if and only if some $A_i \subset X_j$.*
(b) *$X_i \subset \sum B_j$ if and only if $X_i \subset$ some B_j.*
(c) *$\sum A_i \subset B$ if and only if every $A_i \subset B$.*
(d) *$A \subset \prod B_j$ if and only if $A \subset$ every B_j.*
(e) *$\prod A_i \subset \sum B_j$ if and only if $\prod A_i \subset$ some B_j or some $A_i \subset \sum B_j$.*
(f) *$A_i \subset \sum A_j$, $B_i \supset \prod B_j$ all i.*
(g) *$\sum A_i$ is the least upper bound to $A_1, \cdots, A_n$ under $\subset$.*

Proof. By repetition of (12), (11).

To answer "Is $A = B$?," apply (2) and theorem 1. Conditions (12) or (17) are more convenient in practise than (5). (17*b*) is like the condition that a prime number divide a product.

3. Canonical forms

Having thus found one collection of elements equal to each other, and likewise other collections, we should like to choose as canonical forms one element from each collection.

[4] Cf. Schröder, *Algebra der Logik*, vol. 1, p. 184, or MacNeille, *Partially Ordered Sets*, Trans. Am. Math. Soc., 42 (1937), pp. 416–60, or it may be readily verified directly.

THEOREM 2. *Of all the elements equal in a free lattice, there is one of shortest length, unique except for commutativity and associativity.*

PROOF. We show by induction on $k = L(A) + L(B)$ that if $A \cong B$, $A \not\equiv B$, then $C \exists: A \cong B \cong C$, and $L(C) < L(A)$ or else $L(C) < L(B)$. Then if the theorem were false, say A and B were alleged to be both of the shortest length, we could find a still shorter element unless $A \equiv B$! It is true vacuously for $k = 2$. Induction:

Case 1. $\sum A_i \cong \sum B_i$. Here, but not in the previous section, we assume A_i and B_i not themselves joins. $\sum A_i \subset \sum B_i$ by (9). $\therefore A_i \equiv \prod_j a_j^i \subset \sum B_n$ (all i) by (17c). $\therefore$ by (17e), for any i, (i) some $a_p^i \subset \sum B_j$ or (ii) $A_i \subset$ some B_r. Note that (ii) holds if $A_i \equiv X_s$, by (17b). Similarly, for any i, (iii) some $b_p^i \subset \sum A_j$ or (iv) $B_i \subset$ some A_r.

If (i) holds for some i, then $a_p^i \subset \sum B_j \subset \sum A_j$ by (9). $A_j \subset \sum A_n$, $j \neq i$. $\therefore a_p^i \cup \sum_{j \neq i} A_j \subset \sum A_j$. Also $A_i \subset a_p^i$, so $\sum A_j \subset a_p^i \cup \sum_{j \neq i} A_j$; $\therefore a_p^i \cup \sum_{j \neq i} A_j \cong \sum A_j$ and the theorem holds in this case; likewise if (iii) holds for some i.

Otherwise, (ii) and (iv) hold for all i. $\therefore$ for all i, $A_i \subset$ some $B_{f(i)}$, $B_i \subset$ some $A_{g(i)}$ [f, g need not be single-valued]. If $i \exists: g[f(i)] \neq i$, then $A_i \subset B_{f(i)} \subset A_j$ ($j \neq i$), $\sum_{n \neq i} A_n \cong \sum A_n$, and the theorem holds; similarly if $f[g(i)] \neq i$. If not, then $f[g(i)] = i$, $g[f(i)] = i$ for all i. $\therefore A_i \cong B_{f(i)}$, $B_i \cong A_{g(i)}$ all i. But $\sum A_i \not\equiv \sum B_i$ by hypothesis, $\therefore p \exists: A_p \not\equiv B_{f(p)}$ or $B_p \not\equiv A_{g(p)}$, say the former. $\therefore$ by induction, $D \exists: A_p \cong B_{f(p)} \cong D$, $L(D) < L(A_p)$ or $L(D) < L(B_{f(p)})$, say the former. Then $D \cup \sum_{n \neq p} A_n \cong \sum A_n$ and the theorem holds.

Case 2. $\prod A_i \cong \prod B_i$. Dual.

Case 3. $\sum A_i \cong \prod B_j$. $\therefore \sum A_i \subset \prod B_j$; hence every $A_i \subset \prod B_j$ and $\sum A_i \subset$ every B_j. Also $\prod B_j \subset \sum A_i$,

$$\therefore \prod B_j \subset \text{some } A_p \quad \text{or some } B_p \subset \sum A_i.$$

$$A_p \subset \prod B_j \text{ by above.} \qquad \sum A_i \subset B_p \text{ by above.}$$

$$A_p \cong \prod B_j \cong \sum A_i. \qquad \therefore B_p \cong \sum A_i \cong \prod B_j. \quad \text{Q.E.D.}$$

We take this unique shortest form as the canonical form.

(18) COROLLARY 1. $A \equiv \sum A_i \equiv \sum_i (\prod_j a_j^i) \cup \sum_{i \in E} X_i$ (E any subset of $1, \cdots, n$) is canonical if and only if (a) no $a_p^i \subset \sum A_n$ and (b) no $A_i \subset \sum_{n \neq i} A_n$ and (c) every A_i is canonical. Dually for $\prod A_i$.

Corollary 1 follows from the proof of theorem 2. From case 3 we might also require: no $A_p \cong \sum A_n$, but this is included in (b).

If $\sum A_i$ is not canonical, we can find the canonical element equal to it, for if (a) is false, then $a_p^i \cup \sum_{j \neq i} A_j \cong \sum A_j$, if ($b$) is false, then $\sum_{n \neq i} A_n \cong \sum A_n$, if ($c$) is false then replace it by its canonical form; in any case we get a shorter, equal element, to which we can again apply the process.

We can now build up a diagram of the free lattice step by step, starting with

the shortest elements. We could not hope to get it all at once since there are infinitely many distinct elements if there are more than two generators.[5]

(19) COROLLARY 2. If $\sum A_i \cong \sum B_i$, and $\sum A_i$ is canonical, then $(a) B_i \subset \sum A_i$, all i, and (b) given $i, j \ni: A_i \subset B_j$.

PROOF. (a) follows from (12c), and if $A_i \equiv X$, then (b) follows from (12c, b). Otherwise $A_i \equiv \prod_j a_j^i \subset \sum B_i$ by (12c). $\therefore$ by (12e), either $A_i \subset$ some B_j as desired, or some $a_j^i \subset \sum B_i \cong \sum A_i$ and then by (18a) $\sum A_i$ is not canonical contrary to hypothesis.

4. Covering theorems

DEFINITION. A **covers** B if $A > B$ and no $C \ni: A > C > B$.

We present some scattered results on the existence of such pairs of elements. Denote the free lattice of n generators by F_n. In this section (but not previously) we assume n finite.

THEOREM 3. *In the free lattice of n generators, if* $A > (\sum_{i \neq p} X_i) \cap (\sum_{i \neq q} X_i)$, *then* $A = \sum_1^n X_i$ or $\sum_{i \neq p} X_i$ *or* $\sum_{i \neq q} X_i$.

We first prove a

LEMMA. *In* F_n, $A \equiv \sum A_i > \sum_{i \neq p} X_i$ *if and only if given* $i, j \ni: A_j \geqq X_i$.

PROOF OF LEMMA by induction on $L(A)$. Obvious for $L(A) \leqq n$. Suppose $L(A) = k > n$, obviously redundant terms having been omitted; e.g., $X_1 \cup X_1$ contracted to X_1. Also, by (17b), we may suppose no A_i is itself a join. Then some A_i has at least two factors, say $A_1 \equiv \prod a_j$. Then

$$\sum_{i \neq p} X_i < \sum A_i \leqq a_s \cup \sum_{i \neq 1} A_i \qquad \text{(all } s).$$

By induction, given $i, j \ni: j^{\text{th}}$ term $\geqq X_i$. If for some s this term is an A_i, then the lemma holds; otherwise $a_s \geqq X_i$ for all s, and $A_1 \geqq X_i$.

COROLLARY 1. *In* F_n, $\sum A_i = \sum_1^n X_i$ *if and only if given* $i, j \ni: A_j \geqq X_i$. $\sum A_i = \sum_{i \neq p} X_i$ *if and only if given* i $(i \neq p)$, $j \ni: A_j \geqq X_i$, *but no* $A_j \geqq X_p$.

COROLLARY 2. *In* F_n, $\sum_1^n X_i$ *covers* $\sum_{i \neq p} X_i$.

COROLLARY 3. *In* F_n, *every* $A \neq \sum_1^n X_i$ *is* $\leqq \sum_{i \neq p} X_i$ *for some* p. Proved by induction.

The rest of theorem 3 is now proved in much the same way.

COROLLARY 4. *If* F_n, $\sum_{i \neq p} X_i$ *covers* $(\sum_{i \neq p} X_i) \cap (\sum_{i \neq q} X_i)$, *any* q.

COROLLARY 5. *In* F_n, *if* $A > X_p$, *then* $A \geqq X_p \cup (\prod_{i \neq p} X_i)$.

THEOREM 4. $\sum_1^n X_i$ *covers* $\sum_{i \neq p} X_i$ *which in turn covers* $(\sum_{i \neq p} X_i) \cap (\sum_{i \neq q} X_i)$, *and* $X_p \cup (\prod_{i \neq p} X_i)$ *covers* X_p, *in any lattice generated by the* X_i *in which these elements are distinct.*

For any other lattice is a homomorphic image of the free one.

The duals of these results are of course likewise true.

HARVARD UNIVERSITY

[5] Cf. G. Birkhoff, *op. cit.*, p. 451. The distinctness of his elements in the free lattice can also be shown by (12). Note that, taking every sixth element, we get an infinite "chain" of distinct elements.

ANNALS OF MATHEMATICS
Vol. 42, No. 1, January, 1941

CORRECTIONS TO OUR PAPER ON THE EXISTENCE OF MINIMAL SURFACES OF GENERAL CRITICAL TYPES

BY MARSTON MORSE AND C. TOMPKINS

The following corrections should be made in our paper which appeared in volume 40 (1939) of these Annals.

(1). In equation (5.14) interchange $\varphi(\beta)$ and $\psi(\alpha)$.

(2). The argument on page 454 following equation (5.16) should be replaced by the following. The case $\alpha = \beta$ in (5.16) may be discarded. When $\alpha \neq \beta$ (5.16) occurs only when $\psi(\alpha)$ is constant on some interval. In this case the harmonic surface defined by $\psi(\alpha)$ is not a minimal surface (see Radó, loc. cit., p. 75). The transformations used in §6 suffice in this case to prove that $A(\varphi)$ is upper reducible at ψ.

(3). Replace the first two paragraphs on page 460 by the following. A directly conformal 1–1 transformation of the disc $r \leqq 1$ into itself induces a transformation $T(\theta)$ of the circle $r = 1$ into itself. A curve $[p]$ of Z will be said to be *equivalent* to $\varphi(\alpha)$ if for a suitably chosen transformation $T(\theta)$ of the above type, the transformation $\varphi(T(\theta))$ "defines" $[p]$.

(4). In the fourth line following Lemma 6.1 replace "the point $f(p)$ of Ω" by a point $f(p)$ of Ω equivalent to $[p]$ and varying continuously with $[p]$.

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON UNIVERSITY.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

ON THE LUSTERNIK-SCHNIRELMANN CATEGORY

BY RALPH H. FOX

(Received August 10, 1939)

Introduction. In their study of the calculus of variations in the large [36; 37],[1] in particular, of the existence of geodesics on a surface of the topological type of the sphere, Lusternik and Schnirelmann were led [38; 39; 40; 42] to introduce new topological invariants the category and the combinatorial or homology (mod 2) category. The rôle of the homology category was to furnish a medium for the application of homology theory to the calculation of the category of manifolds.

Although the category occupies a unique place in the theory of the calculus of variations in the large—a place which invites further study—it is not to this application that this investigation is directed. My principal objective is to study the relationships between the categories and the standard topological invariants (homology and homotopy groups, homotopy type, etc.). This is a particularly interesting view-point, first noted by Borsuk [14], because the categories seem to have a large measure of independence from these invariants.

The carrying out of this program is greatly facilitated by "dimensionalization" of the categories. It is reasonable to expect that such an analysis of the categories should prove to be a useful tool. This dimensionalization is relatively easy to carry out in the case of homology category and has been implicitly indicated by Eilenberg [20, p. 187]; in the case of the homotopy category it is necessary to define an "n-dimensional homotopy." It is hoped that this dimensionalized homotopy will be of much wider use than in the study of category. There is a strong analogy between the homotopy categories and the homotopy groups, and between the homology categories and the homology groups.

In Chapter 1, I redefine the category, making use of open, instead of the previously used closed, coverings. This is by no means a trivial change. The old definition, as Borsuk [14] has observed, applies to any topological space, while the new definition is not so generally applicable. A little reflection will convince one that those places for which the new definition fails are of little interest in connection with category. Furthermore, the new definition is readily susceptible to the dimensionalization mentioned above. Chapter 1 is written in such a way that the results are valid for the categories defined in Chapters 2 and 3.

In Chapter 2, I define *homotopy in dimension n* and introduce the corresponding *n-dimensional category*. Relations between category, n-dimensional category,

[1] The numbers in brackets refer to the bibliography at the end of the paper.

and various classical invariants are developed. Chapter 3 is devoted to the homology categories.

In Chapter 4, I consider invariants which I call the strong categories. The strong homotopy category was defined implicitly by Borsuk [15]. Just as in the case of the categories, a strong category is defined for each of the relations:

homology in dimension n, homology

homotopy in dimension n, homotopy.

The strong categories exhibit a much greater degree of independence than the categories; for instance, they are not dependent on either the categories, the homotopy type, or deformation retraction. Whether these independencies hold over the class of manifolds is an open and vital question.

Finally, in Chapter 5, I indicate some extensions of the notion of category; in particular a connection between the categories and the multicoherence.

In addition to new results on category many of the previous results have been extended, deepened, and recast.

I have intended this paper to be definitive; I believe that almost all previous results (except applications to calculus of variations and differential geometry) have been included. The bibliography is believed to be complete. Professors Lefschetz and Hurewicz have made many kind suggestions and criticisms. Their inspiration has been invaluable.

I. Homotopy Category

1. Definitions. Let X and M be separable, metric spaces[2] and let f_0 and f_1 be mappings[3] $\epsilon\ M^X$. The mappings f_0 and f_1 are said to be *homotopic in* M, written $f_0 \simeq f_1$ in M, if there is a mapping $f \epsilon M^{X \times [0,1]}$ such that $f(x, 0) = f_0(x)$ and $f(x, 1) = f_1(x)$ for every $x \epsilon X$. Such a mapping f is called a *homotopy between* f_0 and f_1. In general it is important to know in what space a given homotopy occurs, because a mapping space K^X is considered (in an obvious way) as a subspace of M^X whenever $K \subset M$. If there is no danger of confusion the phrase "in M" may be omitted.

When $X \subset M$, X is said to be *deformable in* M *into* $Y \subset M$ if there is a mapping $f_0 \epsilon Y^X$ which is homotopic in M to $f_1 = 1$, the identity[4] mapping of X. A homotopy between f_0 and 1 is called a *deformation of* X *into* Y.

If there is a point $m \epsilon M$ such that X can be deformed in M into m then the set X is said to be *contractible in* M. Thus X is contractible in M when the identity mapping f_1 of M^X is homotopic to a constant[5] mapping f_0 of M^X. A deformation of X into m is called a *contraction of* X *in* M.

A subset A of M will be called a *categorical subset* of M if there is an open

[2] In this paper only separable metric spaces will be considered.

[3] By a *mapping* of M^X is meant a continuous function defined on X with values in M.

[4] An identity mapping of X is the function defined by $f(x) = x$ for every $x \epsilon X$.

[5] A constant mapping f_0 is defined by $f_0(x) = m$ for every $x \epsilon X$. It is sometimes convenient to confuse the mapping f_0 and the point $m \epsilon M$.

set U of M which contains A and is contractible in M. Since any subset of a set contractible in M is itself contractible in M every categorical subset of M is contractible in M; the converse is not true. A covering of X by categorical subsets of M will be called a *categorical covering of X in M*. We shall denote the collection of categorical coverings of X in M by $C_M(X)$.

For any covering σ of some space let us denote by $|\sigma|$ the number of sets of σ. The *category*, $\text{cat}_M X$, of X in M is here[6] defined to be the smallest of the cardinal numbers $|\sigma|$ as σ ranges over $C_M(X)$. A categorical covering σ of X in M will be said to be *minimal* if $|\sigma| = \text{cat}_M X$.

2. The Basis for Abstraction. Many theorems on category do not utilize fully the special character of the relation of homotopy. In the present chapter results of this type will be developed. For this purpose I shall write $\leftrightarrow$ for any relation of equivalence (i.e. one which is symmetric, reflexive and transitive) which has the following properties:

(2.1) If $f_0 \leftrightarrow f_1$ in K then $f_0 \leftrightarrow f_1$ in M for every M which $\supset K$.

(2.2) If ϕ_0 and $\phi_1 \in X^P$ and f_0 and $f_1 \in M^X$ are such that $\phi_0 \leftrightarrow \phi_1$ and $f_0 \leftrightarrow f_1$, then $f_0\phi_0 \leftrightarrow f_1\phi_1$.

(2.3) If A and B are mutually separated,[7] M arcwise connected and f_0 and $f_1 \in M^{A+B}$ such that $f_0 \mid A \leftrightarrow f_1 \mid A$ and $f_0 \mid B \leftrightarrow f_1 \mid B$ then $f_0 \leftrightarrow f_1$.

(2.4) If f_0 and $f_1 \in M^X$ such that $f_0 \leftrightarrow f_1$ and if x_0 and $x_1 \in X$ are such that $f_0(x_0)$ and $f_1(x_0)$ belong to an arc of M then $f_0(x_1)$ and $f_1(x_1)$ also belong to an arc of M.

(2.5) If ϕ_i and $\psi_i \in M_i^{X_i}$ $(i = 1, 2)$ such that $\phi_1 \leftrightarrow \psi_1$ and $\phi_2 \leftrightarrow \psi_2$ then $\phi \leftrightarrow \psi$ where $\phi = (\phi_1, \phi_2)$ and $\psi = (\psi_1, \psi_2) \in M^X$ and $X = X_1 \times X_2$ and $M = M_1 \times M_2$.

The relation h (homotopy) has these properties. Later we shall meet relations h_n (homotopy in dimension n), H (homology) and H_n (homology in dimension n) which also have these properties. In the remainder of this chapter (till §13) the definitions in §1 of deformation, contraction and category are to be read in the sense that the underlying equivalence relation is any predetermined $\leftrightarrow$. In particular, $\leftrightarrow$ may be h, h_n, H or H_n.

Two consequences of (2.2) are sufficiently important to be explicitly pointed out:

(2.6) If f_0 and $f_1 \in M^X$ such that $f_0 \leftrightarrow f_1$ and $A \subset X$ then $f_0 \mid A \leftrightarrow f_1 \mid A$.

(2.7) The relation $\leftrightarrow$ is a topological invariant.

To demonstrate (2.6) choose $P = A \subset X$ and $\phi_0 = \phi_1 =$ identity mapping of A. Then $f_0\phi_0 = f_0 \mid A$ and $f_1\phi_1 = f_1 \mid A$. The proof of (2.7) is trivial.

[6] This definition is applicable only to spaces X, M which have the property that every point of X is a categorical subset of X in M. This restricts the range of the definition from the generality of the definition used by Borsuk [14]. Moreover, even when they both apply, they need not be the same. However, as we shall see later, the definitions coincide in the important case that M is an absolute neighborhood retract, (see §14).

[7] Spaces X and Y are said to be mutually separated if they are disjoint and open in their union $X + Y$.

From now on we assume that $C(M)$ is not vacuous; this condition is satisfied, for instance, when M is locally contractible.

3. Elementary Relations. The function $\text{cat}_M X$ is an increasing function of X and a decreasing function of M in the following sense:

(3.1) *If* $X \subset Y$ *then* $\text{cat}_M X \leqq \text{cat}_M Y$. *If* M *is open in* N *then* $\text{cat}_M X \geqq \text{cat}_N X$.

The proof is obvious.

Let σ and σ' be two coverings. I shall say that σ is a *precise refinement* of σ' if there is a (1, 1) correspondence between the sets of σ and σ' such that every set of σ is contained in the corresponding set of σ'.

(3.2) *There is always a minimal covering belonging to* $C_M(X)$ *which is open. If* X *is closed and* $\text{cat}_M X$ *is finite there is a closed minimal covering* $\epsilon\, C_M(X)$. *If* X *is closed and* M *is a complex*[8] *there is a closed minimal covering* $\epsilon\, C_M(X)$ *whose sets are subcomplex of* M *in a certain subdivision.*[9]

For every categorical covering of X in M is a precise refinement of an open categorical covering. Every finite open categorical covering has a closed categorical precise refinement. If M is a complex it may be subdivided so fine that every simplex is contained in at least one of the sets of a preassigned open categorical covering.

It is easy to verify that

(3.3) *If* X *is compact,* $\text{cat}_M X$ *is finite.*

4. Further Elementary Relations. There is a "triangle inequality" for category:

(4.1) *For any collection* $\{X_\alpha\}$ *of subsets of* M, $\text{cat}_M (\sum X_\alpha) \leqq \sum \text{cat}_M X_\alpha$.

If, for each α the covering σ_α belongs to $C_M(X_\alpha)$ then the covering σ of $\sum X_\alpha$ which consists of all the sets of all the σ_α's is a covering belonging to $C_M(\sum X_\alpha)$. Clearly, $|\sigma| = \sum |\sigma_\alpha|$.

(4.2) *If* M *is arcwise connected and* X *and* Y *are mutually separated*[7] *then* $\text{cat}_M (X + Y) = \max\{\text{cat}_M X, \text{cat}_M Y\}$.

Since M is completely normal, X and Y are contained in disjoint open sets

[8] In this paper complexes are understood to be finite.

[9] But there is no integer k such that there is always a closed minimal covering $\epsilon\, C_M(X)$ whose sets are subcomplexes of M in the k^{th} subdivision. Let $X = M$ be the 2-dimensional Möbius strip mod m, M^2_m, [2, chapters IV, V, VI, Anhang, 12] obtained from a triangulated pseudoprojective space, P^2_m, by removing the interior of a 2-simplex, Q. The category of M is clearly 2. The required number, k, of subdivisions of M must be so large that each 1-simplex of the boundary of Q is subdivided into at least $2m/3$ 1-simplexes. Thus 2^k must be $\geqq 2m/3$ so that k must be $\geqq \log_2 2/3 + \log_2 m$.

This example, kindly shown to me by S. Eilenberg, was constructed by Eilenberg and J. H. C. Whitehead, to answer the following question of H. Hopf: Can one find, for every integer j, a multicoherent complex which is "simplicially" unicoherent in the j^{th} subdivision? The complex M^2_m has this property with j the largest integer in $\log_2 2/3 + \log_2 m$.

It would be interesting to investigate these problems further; especially with the added restriction that the complex M be a manifold.

U and V respectively. Let σ be a categorical covering of X in M by open sets of U and σ' a categorical covering of Y in M by open sets of V. The covering σ'' of X and Y whose sets are unions of pairs of sets, one from σ and one from σ', is open and, by (2.3), belongs to $C_M(X + Y)$. But σ'' contains a subcovering σ''' for which $| \sigma''' | = \max \{| \sigma |, | \sigma' |\}$. Thus $\text{cat}_M (X + Y) \leqq \max \{\text{cat}_M X, \text{cat}_M Y\}$. But, by (3.1), $\text{cat}_M (X + Y) \geqq \max \{\text{cat}_M X, \text{cat}_M Y\}$.

A space M will be said to be a *divisor* of a containing space N if for any space X and mappings f and $g \in M^X$ the homotopy of f and g in N implies their homotopy in M. For example a retract [6] M of N is a divisor of N, by (2.2).

(4.3) *If M is a divisor of N then $\text{cat}_M X \leqq \text{cat}_N X$.*

For a categorical subset of X in N, which is $\subset$ M is then a categorical subset of X in M.

(4.4) *If M_1 and M_2 are mutually separated and $X \subset M_1$ then $\text{cat}_{M_1+M_2} X = \text{cat}_{M_1} X$.*

For M_1 is a retract of $M_1 + M_2$ and is open in $M_1 + M_2$.

(4.5) *If $X = X_1 + X_2$, $M = M_1 + M_2$ and $X_1 \subset M_1$, $X_2 \subset M_2$ where M_1 and M_2 are mutually separated then $\text{cat}_M X = \text{cat}_{M_1} X_1 + \text{cat}_{M_2} X_2$.*

By (2.4), every categorical subset of X in M is a categorical subset of either X_1 or X_2 in M. Hence, $\text{cat}_M X_1 + \text{cat}_M X_2 \leqq \text{cat}_M X$. But, by (4.1), $\text{cat}_M X \leqq \text{cat}_M X_1 + \text{cat}_M X_2$ so that $\text{cat}_M X = \text{cat}_M X_1 + \text{cat}_M X_2$. By (4.4), $\text{cat}_M X_1 = \text{cat}_{M_1} X_1$ and $\text{cat}_M X_2 = \text{cat}_{M_2} X_2$.

If M is locally arcwise connected, so that the components of M are open, then, by an extension of (4.5), $\text{cat}_M X = \sum \text{cat}_{M_\iota} (X \cdot M_\iota)$, the summation extended over the components M_ι of M. Thus no generality is lost in the investigation of $\text{cat}_M X$ if M is supposed connected (hence, arcwise connected). If one is willing to restrict oneself to locally connected X, it follows in the same way from (4.2) that no generality is lost if X is also assumed connected (hence arcwise connected).

5. Categorical sequences. I shall call a finite sequence $\{A_1, A_2, \cdots, A_k = X\}$ of closed subsets of X a *categorical sequence* for X in M if $A_1 \subset A_2 \subset \cdots \subset A_k$, and if $A_1, A_2 - A_1, \cdots, A_k - A_{k-1}$ are categorical subsets of M. The *length* of a categorical sequence $\{A_1, A_2, \cdots, A_k\}$ is k.

THEOREM 5.1. *If M is arcwise connected and $\text{cat}_M X$ is finite then $\text{cat}_M X$ is the minimum of the lengths of the categorical sequences for X in M. Furthermore, if X is finite dimensional, a categorical sequence $\{A_1, A_2, \cdots, A_k\}$ of minimum length can be chosen so that*

$$\dim A_1 < \dim A_2 < \cdots < \dim A_k .$$

First we prove that if $\{A_1, \cdots, A_k\}$ is a categorical sequence for X in M then $\text{cat}_M X \leqq k$. This is obvious for $k = 1$; suppose that it has been proved for $k \leqq r - 1$ and let $\{A_1, A_2, \cdots, A_r\}$ be a categorical sequence for X in M. Since A_1 is, by assumption, categorical in M there is an open set X_1 containing

A_1 which is contractible in M. It is easy to verify that $\{A_2 - X_1, A_3 - X_1, \cdots, A_r - X_1\}$ is a categorical sequence for $X - X_1$ in M of length $r - 1$. By the induction hypothesis, $\mathrm{cat}_M(X - X_1) \leqq r - 1$. Hence, by (4.1) $\mathrm{cat}_M X \leqq r$, completing the induction.

Next we prove that there is a categorical sequence for X in M of length $\leqq \mathrm{cat}_M X$. This is obvious for $\mathrm{cat}_M X = 1$; suppose that it has been proved for $\mathrm{cat}_M X \leqq r - 1$ and let $\{X_1, \cdots, X_r\}$ be an open minimal covering belonging to $C_M(X)$. Denote by F_i the set of points of X which belong to X_j for $j \leqq i$ but not to X_j for $j > i$. Thus the sets F_i are closed in X. Since F_1 and $X - X_1$ are closed disjoint sets of the normal space X, there is an open set G_1 of X such that $F_1 \subset G_1$ and $\bar{G}_1 \cdot (X - X_1) = 0$. If X is finite dimensional, G_1 can be chosen so that $\dim X \cdot (\bar{G}_1 - G_1) < \dim X$.

Suppose we have constructed $j - 1$, open sets $G_1, \cdots, G_{j-1}$ of X such that, for $i \leqq j - 1$,

(5.2) $G_i \supset F_i - (G_1 + \cdots + G_{i-1})$, and

(5.3) $\bar{G}_i \cdot (X - X_i) = 0$,

and, if X is also finite dimensional, $\dim X \cdot (\bar{G}_i - G_i) < \dim X$.

The sets $X - X_j$ and $F_j - \sum_{i<j} G_i$ are closed in X. Since $(G_1 + \cdots + G_{j-2}) + G_{j-1} \supset F_{j-1}$ we have

$$(F_j - \sum_{i<j} G_i) \cdot (X - X_j) \subset (F_j - F_{j-1}) \cdot (X - X_j) \subset X_j \cdot (X - X_j) = 0.$$

Hence there is an open set G_j of X containing $F_j - \sum_{i<j} G_i$ and such that $\bar{G}_j \cdot (X - X_j) = 0$. Thus we construct inductively r open sets $G_1, \cdots, G_r$ which satisfy (5.2) and (5.3) for every $i \leqq r$. If X is finite dimensional, $\dim X \sum_{i \leqq r} (\bar{G}_i - G_i) < \dim X$.

Since $\bar{G}_1 - G_1 \subset X_1 - F_1 \subset X_2 + \cdots + X_r$, it follows, by (5.3), that

$$\sum_{i \leqq r} (\bar{G}_i - \bar{G}_i) \subset X_2 + \cdots + X_r,$$

so that the category of $\sum_{i \leqq r} (\bar{G}_i - G_i)$ in M is $\leqq r - 1$. Hence, by the induction hypothesis, there is a categorical sequence $\{A_1, \cdots, A_{k-1}\}$ for $\sum_{i \leqq r} (\bar{G}_i - G_i)$ in M whose length $k - 1$ is $\leqq r - 1$.

Since $X \cdot \sum_{i \leqq r} (\bar{G}_i - G_i)$ is closed in X, the sets $X \cdot A_j$ are closed in X. In order to show that $\{X \cdot A_1, \cdots, X \cdot A_{k-1}, X\}$ is a categorical sequence for X in M it remains only to show that $X - X \cdot A_{k-1}$ is categorical in M. But $X - X \cdot A_{k-1} = X - X \cdot \sum_{i \leqq r} (\bar{G}_i - G_i)$ is open in X and every component is contained in one of the sets $G_i \subset X_i$. Since each X_i is contractible in M it follows from (3.1) and (4.2) that $X - X \cdot A_{k-1}$ is categorical in M.

If X is finite dimensional the last statement of the theorem follows inductively from the possibility of choosing the G's in such a way that $\dim X \cdot \sum_{i \leqq r} (\bar{G}_i - G_i) < \dim X$.

From theorem 5.1 we can quickly deduce the Lusternik-Schnirelmann-Borsuk theorem, [14, Satz 4; 18; 40, p. 33; 42, p. 132].

(5.4) *If M is arcwise connected and* $\mathrm{cat}_M X$ *is finite then* $\mathrm{cat}_M X \leqq 1 + \dim X$.

If X is not finite dimensional there is nothing to prove. For finite dimensional

X this is an immediate consequence of the existence of a categorical sequence of minimum length whose sets are of increasing dimension.

6. A property of minimal coverings. If M is arcwise connected and the *finite covering* σ *of* X *by open sets of* M *is a minimal categorical covering of* X *in* M *then*

a) *the nerve of* σ *is a simplex and the image of* X *under the Alexandroff mapping*[10] *intersects every open face of this simplex.*

b) *The distributive lattice generated by the sets of* σ, *under the operations of union and intersection, is free.*[11]

Let $\sigma = \{X_1, \cdots, X_k\}$ be an open minimal categorical covering of X in M. Since X is normal there is a closed covering $\{W_1, \cdots, W_k\}$ with $W_i \subset X_i$, (3.2). Let T_i, $i = 1, \cdots, k$, denote the set of points of X which belong to at least $k - i + 1$ of the sets $W_1, \cdots, W_k$. Using (4.2) it is not difficult to see that $\{T_2 - T_1, \cdots, T_k - T_1\}$ is a categorical sequence for $X - T_1$ in M so that $\operatorname{cat}_M (X - T_1) \leqq k - 1$. Since $\operatorname{cat}_M X = k$ by hypothesis, $T_1 \neq 0$. Hence $X_1 \cdot X_2 \cdots X_k \neq 0$, i.e. the nerve of σ is a simplex (of dimension $k - 1$).

The inverse images of the open faces of this simplex under the Alexandroff mapping are the sets $X_{i_1} \cdot X_{i_2} \cdots X_{i_j} - \sum' X_i$, $i \leqq j \leqq k$, where the summation extends over $i \neq i_1, i_2, \cdots, i_j$. It is to be shown that no such set is vacuous. Suppose, for example, that $X_1 \cdots X_j \subset X_{j+1} + \cdots + X_k$. Choose a closed covering $\{W_1, \cdots, W_k\}$ which is a precise refinement of σ, with $W_i \subset X_i$.

The j closed sets $W_i - (X_{j+1} + \cdots + X_k)$, $i = 1, \cdots, j$, have a vacuous intersection; hence we may choose open sets $U_1, \cdots, U_j$, satisfying

$$U_1 \cdot U_2 \cdots U_k = 0 \qquad \text{and} \qquad W_i - (X_{j+1} + \cdots + X_k) \subset U_i \subset X_i .$$

Let $\sigma' = \{U_1, U_2, \cdots, U_j, X_{j+1}, \cdots, X_k\}$. Then $|\sigma'| = k = \operatorname{cat}_M X$, so that σ' is a minimal covering of $C_M(X)$. On the other hand $U_1 \cdot U_2 \cdots U_k = 0$, so that the nerve is not a simplex. Hence the set $X_1 \cdots X_j - (X_{j+1} + \cdots + X_k)$ is not vacuous.

To prove the second statement it is sufficient to show that a distributive lattice generated by elements $X_1, \cdots, X_k$ is free if no relation of the form $X_1 \cdots X_j + X_{j+1} + \cdots + X_k = X_{j+1} + \cdots + X_k$ holds.[12]

[10] By the Alexandroff mapping I mean here the one defined, say, in [30, p. 93].

[11] For the definition of distributive lattice the reader is referred to [5]. A distributive lattice is free if the only relations are those implied by the axioms for a distributive lattice.

[12] A relation of the lattice is of the form

$$\sum_i \prod_j X_{m_{ij}} = \sum_i \prod_j X_{n_{ij}}; \qquad m_{ij} \neq m_{ik}; \qquad n_{ij} \neq n_{ik} .$$

Let $X_1 \cdots X_t$ be any product of shortest length in this relation and let $X_{t+1}, \cdots, X_k$ be the rest of the elements. Then, adding $X_{t+1} + \cdots + X_k$ to both sides, $X_1 \cdots X_t + X_{t+1} + \cdots + X_k = X_{t+1} + \cdots + X_k$, since every product, with the single exception of $X_1 \cdots X_t$, contains at least one of the elements $X_{t+1}, \cdots, X_k$.

7. Deformation. Now I shall show that deformation can not lower the category. More precisely:

THEOREM 7.1. *If X is open in M and can be deformed in M into Y then* $\operatorname{cat}_M X \leqq \operatorname{cat}_M Y$.

By hypothesis there is a mapping $f_0 \epsilon Y^X$ such that $f_0 \leftrightarrow 1 \mid X$ in M. Let $\sigma = \{Y_i\}$ be an open contractible covering of Y in M, so that the covering $f_0^{-1}(\sigma) = \{f_0^{-1}(Y_i)\} = \{X_i\}$ is open in X and hence in M. Since $1 \mid Y_i \leftrightarrow$ constant, $1 \mid X_i \leftrightarrow f_0 \mid X_i \leftrightarrow$ constant. Thus X_i is contractible so that $f_0^{-1}(\sigma) \epsilon C_M(X)$.

A property is said to be *inductive* [2, II, Anhang, 1] if, whenever each of a decreasing sequence of compact sets has the property, their intersection also has the property.

(7.2) The *property* $\operatorname{cat}_M X = n$ *for fixed M and compact X is inductive.*

To prove this it is sufficient, in view of (3.1), to show that *if $\{X_i\}$ is a sequence of subsets of M such that the closure of $\sum X_i$ is compact then there is an integer i_0 such that* $\operatorname{cat}_M X_i \leqq \operatorname{cat}_M X$ *for every $i \geqq i_0$, where*[13] $X = \overline{\lim} \{X_i\}$.

Since the closure of $\sum X_i$ is compact, X is not vacuous. There is an open set U ($= \sum X_j$ for some open minimal covering $\epsilon C_M(X)$) containing X such that $\operatorname{cat}_M U = \operatorname{cat}_M X$. Since the closure of $\sum X_i$ is compact, U contains almost all X_i. Thus there is an integer i_0 such that $X_i \subset U$ when $i \geqq i_0$. Hence, by (3.1), $\operatorname{cat}_M X_i \leqq \operatorname{cat}_M U$.

A set $A \subset M$ will be said to be *essential* in M if no neighborhood of A can be deformed in M into a proper closed subset of A.

From (7.2) quickly follows:

(7.3) *If X is compact, there is a closed subset A of X which is essential in M and whose category in M* $= \operatorname{cat}_M X$.

By (7.2) and the irreducibility principle [2, II, Anhang, 1] there is a closed subset A of X such that $\operatorname{cat}_M A = \operatorname{cat}_M X$, but $\operatorname{cat}_M B < \operatorname{cat}_M X$ for every proper closed subset B of A. This set A is essential in M, for, if it were not, a neighborhood U could be deformed in M into a proper closed subset B of A. By (3.1) and theorem 7.1 it would follow that $\operatorname{cat}_M A \leqq \operatorname{cat}_M U \leqq \operatorname{cat}_M B$ which is in contradiction with the construction of A, since B is compact.

A closed set X will be said to be *essential in M in dimension n* if X has a closed n-dimensional subset which is essential in M; X will be said to be *essential in M in exactly $r + 1$ dimensions* if there is a sequence of integers $0 = n_0 < \cdots < n_r$ such that X is essential in M in the dimensions $n_0, \cdots, n_r$ and only these.

We can now prove the following refinement of (5.4):

THEOREM 7.4. *If M is arcwise connected and the compact set X is essential in M in exactly $r + 1$ dimensions then* $\operatorname{cat}_M X \leqq r + 1$.

The statement is true for $r = 0$ by (7.3), (3.3), and (5.4).

Suppose the theorem has been proved for $r < m$ and consider a compact X

[13] A point $x \epsilon M$ belongs to $\overline{\lim} \{X_i\}$ if every neighborhood of x intersects an infinite number of the sets X_i.

which is essential in M in exactly $m + 1$ dimensions, $0 = n_0 < \cdots < n_m$. By (7.3) there is a compact set $A \subset X$ which is essential in M and such that $\text{cat}_M A = \text{cat}_M X$. Since X is inessential in M in the dimensions $> n_m$, $\dim A = k \leqq n_m$. Since A is finite dimensional there is a categorical sequence $\{A_1, \cdots, A_k\}$ for A in M of length k such that $\dim A_1 < \cdots < \dim A_k$ (theorem 5.1). Since $\dim A_{k-1} < \dim A \leqq n_m$, the compactum A_{k-1} is essential in at most m dimensions. Hence, by the induction hypothesis, $\text{cat}_M A_{k-1} \leqq m$. By (4.1), we have $\text{cat}_M A \leqq \text{cat}_M A_{k-1} + \text{cat}_M (A_k - A_{k-1}) \leqq m + 1$, completing the induction.

Further discussion of this refinement of (5.4) will follow in §18.

8. Absolute category. Of particular interest is the (absolute) category cat $M = \text{cat}_M$ of M. From (3.1) and (4.1) follows

(8.1) *If M and N are open in $M + N$ then* $\text{cat}(M + N) \leqq \text{cat } M + \text{cat } N$.

From (3.1) and (7.1) follows

(8.2) *If M is open in N, and N can be deformed in itself into M, then* $\text{cat } M \geqq \text{cat } N$.

From (3.1) and (4.3) follows

(8.3) *If M is a divisor of N then* $\text{cat } M \leqq \text{cat } N$.

For simplicity of statement I shall restrict myself in the three succeeding theorems to absolute category. Removal of this restriction is not difficult.

9. Product spaces. For the category of a product space the following inequality has been proved by Bassi [4]:

THEOREM 9. *If $M = M_1 \times M_2$ is arcwise connected and* $\text{cat } M_1$ *and* $\text{cat } M_2$ *are finite, then*

$$\max \{\text{cat } M_1, \text{cat } M_2\} \leqq \text{cat } M \leqq \text{cat } M_1 + \text{cat } M_2 - 1$$

The first inequality is an immediate consequence of (8.3) since $M_1 \times x_2$ and $x_1 \times M_2$ where (x_1, x_2), denotes a point of M, are retracts of M and are homeomorphs of M_1 and M_2 respectively.

To prove the second inequality let $\text{cat } M_1 = m$ and $\text{cat } M_2 = n$. There exists a categorical sequence $\{A_1, \cdots, A_m\}$ for M_1 in M_1 of length m; likewise a categorical sequence $\{B_1, \cdots, B_n\}$ for M_2 can be found. Suppose $m \leqq n$ and define

$$C_k = \sum_{i+j=k+1} A_i \times B_j; \quad k = 1, \cdots, m + n - 1.$$

It remains only to show that the closed sets $C_1, \cdots, C_{m+n-1} = M$ form a categorical sequence for M. One has only to verify that the sets $C_{k+1} - C_k$, $k = 1, \cdots, m + n - 2$ are categorical. But, writing $A_0 = 0$, $B_0 = 0$ for convenience,

$$C_{k+1} - C_k = \sum_{i+j=k+2} (A_i - A_{i-1}) \times (B_j - B_{j-1}), \qquad k = 1, \cdots, m + n - 2.$$

From (2.5) it follows that $(A_i - A_{i-1}) \times (B_j - B_{j-1})$ is a categorical subset of M. Furthermore, if $i < i'$, $j > j'$,

$$\overline{[(A_i - A_{i-1}) \times (B_j - B_{j-1})]} \cdot [(A_{i'} - A_{i'-1}) \times (B_{j'} - B_{j'-1})]$$
$$\subset \overline{(A_i - A_{i-1})} \cdot (A_{i'} - A_{i'-1}) \times M_2$$
$$\subset A_i \cdot (A_{i'} - A_i) \times M_2 = 0$$

and symmetrically

$$[(A_i - A_{i-1}) \times (B_j - B_{j-1})] \cdot \overline{[(A_{j'} - A_{j'-1})(B_{j'} - B_{j'-1})]}$$
$$\subset M_1 \times (B_j - B_{j'}) \cdot B_{j'} = 0.$$

Thus $C_{k+1} - C_k$ is the union of mutually separated categorical subsets. Hence, by theorem 4.2, $C_{k+1} - C_k$ is a categorical set.

10. Homotopy type. Two arcwise connected spaces, M and N, are said to have the same homotopy type [27] if there are mappings $f \in N^M$ and $g \in M^N$ such that $gf \in M^M$ is homotopic to the identity and $fg \in N^N$ is homotopic to the identity.

(10.1) *If there are mappings $f \in N^M$ and $g \in M^N$ such that $gf \in M^M$ is homotopic to the identity then* cat $M \leqq$ cat N.

Let $\sigma = \{Y_i\}$ be an open contractible covering of N and let $X_i = f^{-1}(Y_i)$ so that $f^{-1}(\sigma) = \{X_i\}$ is an open covering of M. Since $g \mid Y_i$ is homotopic to a constant, $gf \mid X_i$ is homotopic to a constant. But $gf \mid X_i$ is homotopic to the identity mapping of M^{X_i}. Hence $f^{-1}(\sigma)$ is contractible.

(10.2) *The absolute category is an invariant of the homotopy type; i.e. if X and Y have the same homotopy type* cat X = cat Y.

11. A characterization of category. It has been shown that when M is a Hausdorff space for which $C(M)$ is not vacuous the category has the following properties:

(i) If X is a point, $\text{cat}_M X = 1$

(ii) If $X \supset Y$ then $\text{cat}_M X \geqq \text{cat}_M Y$, (3.1).

(iii) $\text{cat}_M (\sum X_\alpha) \leqq \sum \text{cat}_M X_\alpha$, (4.1).

(iv) If X is open and can be deformed in M into Y then $\text{cat}_M X \leqq \text{cat}_M Y$, (theorem 7.1).

It will now be shown that these four properties characterize, in a certain sense, the set function $\text{cat}_M X$. [42; 40].

(11) *The set of positive-integer valued functions $\lambda(X)$, defined for subsets of M and satisfying*

(i) *if X is a point then $\lambda(X) = 1$,*

(ii) *if $X \supset Y$ then $\lambda(X) \geqq \lambda(Y)$,*

(iii) $\lambda(\sum X_\alpha) \leqq \sum \lambda(X_\alpha)$,

(iv) *if X is open and can be deformed in M into Y then $\lambda(X) \leqq \lambda(Y)$,*

is partially ordered by the rule: $\lambda_1 \leqq \lambda_2$ *when* $\lambda_1(X) \leqq \lambda_2(X)$ *for every* X; *the category*, $\operatorname{cat}_M X$, *is the largest element of this partially ordered set.*

Let $\sigma = \{X_\alpha\}$ be an open minimal contractible covering of X in M. By (iv) and (i), $\lambda(X_\alpha) = 1$ for each α. By (ii) and (iii) $\lambda(X) \leqq \lambda(\sum X_\alpha) \leqq \sum \lambda(X_\alpha) = |\sigma| = \operatorname{cat}_M X$.

12. Minima of set functions. I conclude this chapter with a theorem which may be considered as the topological part of the Lusternik-Schnirelmann theorem on category and calculus of variations [38; 39; 40]. Let g be a real valued set function defined for the subsets of M and satisfying

(i*) if $X \supset Y$ then $g(X) \geqq g(Y)$.

Denote by $\mathfrak{M}^n$ the collection of sets X for which $\operatorname{cat}_M X \geqq n$ and let $c_n = \inf_{X \in \mathfrak{M}^n} g(X)$. A set X of $\mathfrak{M}^n$ will be said to be minimal (relative to $\mathfrak{M}^n$ and g) if $g(X) = c_n$.

THEOREM 12. *If* $c_m = c_n = c$ $(m < n)$, *if there exists at least one closed minimal set (relative to* $\mathfrak{M}^n$), *and if* D *is a closed set whose category in* M *is* $\leqq n - m$ *then there exists a closed minimal set (relative to* $\mathfrak{M}^m$) *disjoint to* D.[14]

There is an open set $U \supset D$ such that $\operatorname{cat}_M U = \operatorname{cat}_M D$. By assumption there is a closed minimal set X relative to $\mathfrak{M}^n$. The closed set $Y = X \cdot (M - U)$ is, by construction, disjoint to D. It remains only to show that Y is a minimal set relative to $\mathfrak{M}^n$.

Since $X \in \mathfrak{M}^n$ the category of X in M is $\geqq n$. But $X \subset Y + U$ so that, by (ii) and (iii) of §11, $\operatorname{cat}_M X \leqq \operatorname{cat}_M (Y + U) \leqq \operatorname{cat}_M Y + \operatorname{cat}_M U \leqq \operatorname{cat}_M Y + (n - m)$. Hence $\operatorname{cat}_M Y \geqq m$ so that Y belongs to $\mathfrak{M}^m$.

Since Y belongs to $\mathfrak{M}^m$ it follows that $g(Y) \geqq c$. But, since $Y \subset X$, it follows from (*i**) that $g(Y) \leqq g(X) = c$. Therefore $g(Y) = c$.

Since Y belongs to $\mathfrak{M}^m$ and $g(Y) = c_m$, it is a minimal set relative to $\mathfrak{M}^m$.

II. THE n-DIMENSIONAL CATEGORY

13. Homotopy in dimension n. Among the absolute neighborhood retracts [7] the contractible spaces are characterized [25] by the vanishing of all their homotopy groups. This suggests the possibility of characterizing in an analo-

[14] The proof of this theorem uses only properties (ii) and (iii) of §11 and the existence, for any $X \subset M$, of an open neighborhood of the same category; it is therefore valid if cat_M is replaced by a positive-integer valued set function having these properties. The application of this theorem to the calculus of variations is the following: Let M be a compact, connected, finite dimensional Reimannian manifold, f a function on M of class C''. Let $g(X) = \sup_{x \in X} f(x)$ and let D be the set of points x of M where all the partial derivatives of f vanish, i.e. D is the set of stationary points of f. A theorem of Lusternik and Schnirelmann [36; 38; 40, p. 22] states that D intersects every closed minimal set. It follows that if $c_m = c_n$, and if there is a closed minimal set relative to $\mathfrak{M}^n$ and g, then $\operatorname{cat}_M D > n - m$. From this and the above mentioned theorem of Lusternik and Schnirelmann it follows, in particular, that every function of class C'' on M has at least cat M stationary points.

gous way the subsets of an absolute neighborhood retract which are contractible in it. On investigation we find the rather surprising result:

In order that a subcomplex A of a complex M be contractible in M it is not sufficient that every continuous sphere $f \in M^{S_n}$, for which $f(S_n) \subset A$, be homotopic in M to a constant.

Let A be a 2-dimensional torus and M a complex obtained from A by addition of two 2-cells which span a meridian and an equator respectively, there being no other identifications. Every continuous n-sphere $f \in M^{S_n}$, for which $f(S_n) \subset A$, is homotopic in M to a constant. This is clear for $n = 1$ because M has a vanishing fundamental group. For $n > 1$ it is a consequence of the fact that A is an aspherical [28] space. However, A is not contractible in M since there is a 2-cycle in A which does not bound in M.

Thus the contractibility of a subset A of an absolute neighborhood retract M can not be determined by continuous spheres alone. We shall see that a characterization may be given in terms of continuous complexes.

However, contractibility is homotopy of a very special kind. The characterization of contractibility by means of continuous complexes may be extended to a characterization of homotopy. This leads to the important notion of homotopy in dimension n:

Mappings ϕ and ψ of M^X will be said to be *homotopic in dimension n* or *n-homotopic* if for every continuous n-dimensional complex $f \in X^P$ the continuous complexes ϕf and $\psi f \in M^P$ are homotopic. If X is of uniform class LC^n then a necessary and sufficient condition for ϕ and ψ to be n-homotopic is that $\phi f \simeq \psi f$ for every n-dimensional compactum K and mapping $f \in X^K$. The proof is along the lines of [27, §4] using [31, Theorem 5]. I shall write h for homotopy and h_n for n-homotopy.

Observe that homotopic mappings are homotopic in every dimension and that homotopy in dimension n implies homotopy in every dimension $\leqq n$.

The characterization of homotopy mentioned above is the following:

THEOREM 13. *Let A be a closed subset of an absolute neighborhood retract X and let ϕ and ψ be mappings of M^X. If there is a neighborhood U of A such that $\phi \mid U$ and $\psi \mid U$ are homotopic in every dimension $< 1 + \dim X$ then $\phi \mid A$ and $\psi \mid A$ are homotopic.*

For every $\epsilon > 0$ there is a continuous complex $f \in X^P$ (where $\dim P \leqq n$ if X is n dimensional) and a mapping $g \in P^X$ such that the mapping $fg \in X^X$ is homotopic to the identity and $d(x, fg(x)) < \epsilon$ for every $x \in X$, [35]. Choose $\epsilon < d(A, X - U)$ so that $fg(A) \subset U$.

By hypothesis and (2.2), the mappings $\phi f \mid f^{-1}(U)$ and $\psi f \mid f^{-1}(U)$ of $f^{-1}(U)$ into M are homotopic. Hence the mappings $\phi fg \mid A$ and $\psi fg \mid A$ of M^A are homotopic. But $\phi fg \mid A$ is homotopic to $\phi \mid A$ and $\psi fg \mid A$ is homotopic to $\psi \mid A$. Hence $\phi \mid A$ and $\psi \mid A$ are homotopic.

14. A subset A of M will be said to be *h_n-deformable in M* into B if there is a mapping $h_0 \in M^A$, with $h_0(A) \subset B$, which is n-homotopic to the identity

mapping of M^A. I shall say that A is *h_n-contractible in M* if there is a point $m \in M$ such that A can be deformed in M in dimension n into m.

(14.1) *A closed subset A of an absolute neighborhood retract M is contractible in M if and only if there is a neighborhood of A which is contractible in M in every dimension $< 1 + \dim M$.*

The first part is a specialization of theorem 13. The second part follows from

(14.2) *If M is an absolute neighborhood retract in the weak sense*[15] *then a closed subset A is a categorical subset in M if and only if A is contractible in M* [14, theorem 3].

If A is contractible in M there is a point $m \in M$ and a mapping $h \in M^{A \times [0,1]}$ such that $h(x, 0) = m$ and $h(x, 1) = x$ for every $x \in A$. Define a mapping $h' \in M^Q$ where $Q = M \times [0] + A \times [0, 1] + M \times [1]$ by

$$h'(x, 0) = m \text{ for } x \in M$$

$$h'(x, t) = h(x, t) \text{ for } (x, t) \in A \times [0, 1]$$

$$h'(x, 1) = x \text{ for } x \in M$$

Since M is an absolute neighborhood retract in the weak sense and Q is a closed subset of $M \times [0, 1]$, the mapping h' may be extended [31, p. 276 remark 3] to a mapping $h'' \in M^G$ where G is a neighborhood of Q in $M \times [0, 1]$. Let U be an open neighborhood of A in M such that $U \times [0, 1] \subset G$. It is clear that $h'' \mid U \times [0, 1]$ is a contraction of U in M. Hence A is a categorical subset of M.

15. A subset A of M will be called an *h_n-categorical subset* of M if there is an open set U of M which contains A and is h_n-contractible in M. Clearly every h_n-categorical subset is h_n-contractible in M. In contrast to (14.2), without local assumptions on the closed set A, its h_n-contractibility in M does not imply that it is h_n-categorical in M, even with the strongest (non-trivial) assumptions on M. An example to keep in mind is the following: M is a circular ring obtained from the rectangle $|x| \leqq 2/\pi$, $|y| \leqq 1$ in the Cartesian plane by identifying the points $(-2/\pi, y)$ and $(2/\pi, y)$, A is the image under this identification of the closure of the curve $y = \cos 1/x$, $|x| \leqq 2/\pi$ and $n = 1$. In fact it is the existence of such examples which necessitates the introduction of the notion of categorical set.

A covering of X by h_n-categorical subsets of M will be called an *h_n-categorical covering of X in M*. We shall denote the collection of such coverings by $h_nC_M(X)$. The *n-dimensional* (homotopy) *category*, $h_n-\text{cat}_M X$, of X in M is defined to be the smallest of the cardinal numbers $|\sigma|$ as σ ranges over

[15] I call a separable metric space an absolute neighborhood retract in the weak sense if it is a retract of every separable metric containing space in which it is closed. Cf. [31, p. 270 footnote (1)].

$h_n C_M(X)$. A covering σ of $h_n C_M(X)$ will be said to be *minimal* if $|\sigma| = h.\operatorname{cat}_M X$.

The results of chapter 1 have been so worded that they apply to the n-dimensional category. One needs only to substitute homotopy in dimension n for $\leftrightarrow$ and make the implied changes in the succeeding definitions. To see that this is so, it is sufficient to demonstrate that homotopy in dimension n is symmetric, reflexive, and transitive, and has properties (2.1) $\cdots$ (2.5). With the possible exception of (2.5), these verifications are easy.

Let ϕ_1 and $\psi_1 \in M_1^{X_1}$ be homotopic in dimension n and also ϕ_2 and $\psi \in M_2^{X_2}$ be n-homotopic. It is to be shown that $\phi = (\phi_1, \phi_2)$ and $\psi = (\psi_1, \psi_2) \in M^X$, where $X = X_1 \times X_2$ and $M = M_1 \times M_2$, are homotopic in dimension n. Let $f \in X^P$ be a continuous n dimensional complex. Clearly $\phi f = (\phi_1\pi f, \phi_2\pi f)$ and $\psi f = (\psi_1\pi_1 f, \psi_2\pi_2 f)$, where π_1 and π_2 denote the projections of X into X_1 and X_2 respectively. By hypothesis $\phi_1\pi_1 f$ and $\psi_1\pi_1 f$ are homotopic, likewise $\phi_2\pi_2 f$ and $\psi_2\pi_2 f$ are homotopic. Hence, by property (2.5) of homotopy, ϕf and ψf are homotopic. This completes the verification of (2.5) for n-homotopy.

16. As analogue of (14.2) we have:

(16.1) *If M is of uniform class*[16] *LC^n then a subset A of uniform class LC^{n-1} is an h_n-categorical subset of M if and only if it is h_n-contractible in M.*

Since M is of uniform class LC^n there[17] is an $\epsilon > 0$ such that continuous n-dimensional complexes $g_1, g_2 \in M^P$ are homotopic whenever the distance between them is less than ϵ. Since A is of uniform class LC^{n-1} there is an $\eta > 0$ such that every partial realization of a continuous n-dimensional complex in A of mesh $< \eta$ can be completed in A to a full realization of mesh $< \epsilon/3$.

Let U be an $\eta/3$ neighborhood of A and let $f \in U^P$ be a continuous n-dimensional complex. Subdivide P so fine that the image of every simplex is of diameter $< \eta/3$. For each vertex p of P choose a point x of A such that $d(f(p), x) < \eta/3$. Define $f'(p) = x$ for every vertex p. Thus f' is a mapping of the 0-dimensional framework of P into A. The mesh of this partial realization is $< \eta$, hence f' can be completed to a full realization f'' of mesh $< \epsilon/3$. Thus $f'' \in A^P$ and $d(f, f'') < \frac{\eta}{3} + \frac{\eta}{3} + \frac{\epsilon}{3} < \epsilon$. Hence f and f'' are homotopic in M. But, by hypothesis, f'' is homotopic in M to a constant. Hence f is homotopic in M to a constant. Thus U is h_n-contractible in M.

(16.2) *If M is connected and of uniform class LC^1 and X is compact then there is a minimal covering of $h_1 C_M(X)$ whose sets are continuous curves of dimension $\leqq \max\{1, \dim X\}$.*[18]

---

[16] A metric space (not necessarily compact) will be said to be of uniform class LC^n if for every $\epsilon > 0$ there is a $\delta > 0$ such that every continuous k-sphere ($k \leqq n$) of diameter $< \delta$ can be extended to a continuous $(k + 1)$-cell of diameter $< \epsilon$. For compact spaces, LC^n is identical with uniform LC^n.

[17] [34, theorem 1]. The assumption of compactness is unnecessary. In the statement K_p should be at most $(p + 1)$-dimensional.

[18] cf. [14, theorem 5]. A minimal covering of $h_n C_M(X)$ whose sets are continuous of class

Let σ be an open minimal covering of $h_1C_M(X)$ and let σ' be a precise closed refinement. Let U be a set of σ and A the set of σ' contained in U. There is a compact set B of class LC^0, which contains $A \cdot X$ and is contained in U, such that $\dim (B - A \cdot X) \leqq 1$, [14, Lemma 5; 44]. The set B has a finite number, k, of components, and, since M is arcwise connected, B may be enlarged to a continuous curve by the addition of $k - 1$ spanning arcs. Let C denote such a continuous curve constructed inductively by adding spanning arcs one at a time in such a way that at each stage the number of components is decreased by one. This set C is h_1-contractible in M because every continuous 1-dimensional complex in a component of B is homotopic to a constant in M and such a homotopy $h(x, t)$ can be found for which $h(x_0, t) = h(x_0, 0)$ for a preassigned x_0 of the antecedent and every $t \in [0, 1]$. By its construction, $\dim C \leqq \max \{1, \dim X\}$. By (16.1), C is an h_1-categorical subset of M. The collection of C's is the required minimal covering.

17. Category and n-dimensional category. It is quite clear that

(17.1) *$h_n \operatorname{cat}_M X \leqq \operatorname{cat}_M X$ and, if $k \leqq n$, $h_k \operatorname{cat}_M X \leqq h_n \operatorname{cat}_M X$.*

The question: "Under what conditions does equality hold?" receives a partial answer in

(17.2) *If X is a closed subset of an n-dimensional absolute neighborhood retract, M, then $h_n \operatorname{cat}_M X = \operatorname{cat}_M X$.*

Let σ be an open covering of $h_nC_M(X)$ and let $\sigma' \in h_nC_M(X)$ be a precise closed refinement (3.2). By (14.1), each set of σ' is contractible in M. Hence, by (14.2), σ' is a categorical covering of X in M so that $\operatorname{cat}_M X \leqq |\sigma'| = h_n \operatorname{cat}_M X$.

Another partial answer is given by

(17.3) *Let X be a finite dimensional closed subset of an aspherical absolute neighborhood retract M. Then $\operatorname{cat}_M X = h_1 \operatorname{cat}_M X$.*

By (16.2), there is a minimal covering σ of $h_1C_M(X)$ whose sets are finite dimensional continuous curves. Since each set of σ is h_1-contractible in the aspherical absolute neighborhood retract M, each set is contractible in M [28, Principal theorem]. Since M is an absolute neighborhood retract it follows from (14.2) that each set of σ is categorical in M. Thus $\operatorname{cat}_M X \leqq |\sigma| = h_1 \operatorname{cat}_M X$.

18. An upper bound for the category. According to Borsuk [11, p. 254] a compact space M is called a homotopy membrane in dimension k if every closed at most k-dimensional subset is contractible in M. From (14.2) it follows that a homotopy membrane in dimension k which is also an absolute neighborhood retract is inessential in all dimensions $\leqq k$. Hence, from theorem 7.1 we have

---

LC^k ($k \geqq n - 1$) can be chosen if X is compact and M is connected and of uniform class LC^n and has the following property: For every compact set A and open set $U \supset A$ there is a compact set B of class LC^k such that $A \subset B \subset U$. It is a little difficult to see what this condition means for $k > 0$.

(18.1) *If the m-dimensional absolute neighborhood retract M is a homotopy membrane in dimension* $k(\leqq m)$ *then* cat $M \leqq m - k + 1$.

Since the vanishing of the first k homotopy groups of a finite dimensional absolute neighborhood retract M implies that M is a homotopy membrane in dimension k, [25 Satz 5] and, since the first k homotopy groups of an absolute neighborhood retract M vanish if and only if M is simply connected and acyclic in the dimensions $\leqq k$, [26],

THEOREM 18.2. *If the m-dimensional absolute neighborhood retract M is simply connected and acyclic in the dimension* $\leqq k(\leqq m)$ *then* cat $M \leqq m - k + 1$.

Theorem 18.2 differs from (18.1) only if M is not a complex, for a complex is a homotopy membrane in dimension k if *and only if* it is simply connected and acyclic in the first k dimensions [26, 3′]. The simply connected absolute neighborhood retract $B(k, m)$, $2 \leqq k \leqq m$, of Borsuk [11, p. 256] is a homotopy membrane in dimension $m - 1$ but is acyclic only in the dimensions $\leqq k - 1$. Thus cat $B(k, m) \leqq m - k + 2$ by theorem 18.2 but cat $B(k, m) \leqq 2$ by (18.1).

In the bounds given by theorem 7.4, (18.1), and theorem 18.2 the equality sign need not hold, even when M is a manifold. In fact let $M = S_1 \times S_2$. According to corollary 20.3, cat $M = 3$. But M is essential in dimensions 0, 1, 2, and 3 so that cat $M \leqq 4$ is the best bound obtainable from the above mentioned theorems.

A similar upper bound for the category follows from an unpublished result of W. Hurewicz: If M is a simply connected m-dimensional complex and the last k Betti groups, with the real number mod. 1 for coefficient domain, vanish, then M can be deformed into its $(m - k)$ dimensional framework. Hence from theorem 7.1 and (5.4) we conclude:

(18.3) *If M is a simply connected m-dimensional complex and* $\beta_i(M) = 0$ *for* $i = m - k + 1, \cdots, m$ *(real numbers* mod 1) *then* cat $M \leqq m - k + 1$.

In a similar vein:

(18.4) *If M is a connected aspherical absolute neighborhood retract whose fundamental group is free with a finite number of generators then* cat $M \leqq 2$.

For, under the conditions of the theorem, there is a compact at most 1-dimensional set which is a deformation retract of M, [21]. As above, the statement follows from theorem 7.1 and (5.4).

19. We note some further results connected with the notion of homotopy type.

(19.1) *If M is a finite dimensional aspherical absolute neighborhood retract with Abelian fundamental group then* cat $M = h_1$ cat $M = 1 + b_1(M)$, *where* $b_1(M)$ *denotes the* 1-*dimensional Betti number of M*.

For, under the conditions of the theorem, M belongs to the homotopy type of the $b_1(M)$-dimensional torus [28]. The result follows from (10.2) and corollary 20.3.

(19.2) *If M is an absolute neighborhood retract and if for every* $\epsilon > 0$ *there is an* ϵ-*mapping of M into a metric space M′ then* cat $M \geqq$ cat M'.

This follows from (10.2) and a theorem of Eilenberg [22].

20. Product spaces. From theorem 9 it follows that, if $M = M_1 \times \cdots \times M_k$ is arcwise connected,

$$\max \{\text{cat } M_i\} \leqq \text{cat } M \leqq 1 + \sum_{i=1}^{k} (\text{cat } M_i - 1), \text{ and} \tag{20.1}$$

$$\max \{h_n \text{ cat } M_i\} \leqq h_n \text{ cat } M \leqq 1 + \sum_{i=1}^{k} (h_n \text{ cat } M_i - 1). \tag{20.2}$$

The simplest examples show that the lower bounds in (20.1) and (20.2) are not always attained; I shall now show that it is the same with the upper bounds in (20.1) and, for $n > 1$, in (20.2).

Let $k = 2$, $M_i =$ the three dimensional pseudoprojective space $P^3_{m_i}$, $i = 1, 2$, [2, VI Anhang 6, 7, 8, p. 266] where m_1 and m_2 are relatively prime. It is easy to show, by construction[19] of a categorical covering, that cat $P^3_{m_i} = h_2$ cat $P^3_{m_i} = 2$. The complexes $M = M_1 \times M_2$ and $T = x_1 \times M_2 + M_1 \times x_2$, where $(x_1, x_2) \in M$, are simply connected. Moreover the homology groups of M and T are the same in every dimension, and for each r the natural homomorphism of $\beta_r(T)$ into $\beta_r(M)$ is an isomorphism which covers $\beta_r(M)$ [2, VII, §2, 3]. Hence, by an unpublished theorem of Hurewicz, T is a deformation retract of M. Since, obviously, h_2 cat $T =$ cat $T = 2$, it follows from theorem 7.1 that h_2 cat $M =$ cat $M = 2$. Thus the upper bound, 3 in this case, is not attained. (I do not know whether the upper bound is always attained if the further assumption that M be essential is imposed.)

A lower bound of a different type has been given for the homotopy category by Eilenberg [19].

THEOREM 20.3. *If $M = M_1 \times \cdots \times M_k$ is an absolute neighborhood retract and essential then* cat $M_i \geqq 2$ *for* $i = 1, \cdots, k$ *implies that* cat $M \geqq k + 1$.

The proof of this theorem, which I omit, depends on the following lemma on homotopy:

(20.4) *If A is a closed subset of an absolute neighborhood retract M and $h' \in M^{A \times [0,1]}$ is a deformation of A in M then there is a deformation $h \in M^{M \times [0,1]}$ such that $h \mid A \times [0, 1] = h'$.*

I do not know whether the theorem or the lemma retains its validity for n-homotopy.

Notice that the example above shows that the condition that M be essential in theorem 20.3 can not be removed, or even be replaced by the condition that each M_i be essential.

Examination of the proof of theorem 20.3 reveals that the condition that M be essential can be replaced by the apparently weaker condition that M can

---

[19] It is not difficult to verify that the category of a join AB is min {cat A, cat B}. The join A and B of two spaces A and B is the space obtained from the product $A \times B \times [0, 1]$ by identifying $(x, y, 0)$ with x and $(x, y, 1)$ with y for every $x \in A$ and $y \in B$. The pseudo-projective space P^3_m is the join of the 2-dimensional pseudo-projective space P^2_m and a 0-sphere S_0.

not be deformed into M^p, the set of points $x = (x_1, \cdots, x_k)$ of M which have at least one coordinate identical with the corresponding coordinate of a fixed point $p = (p_1, \cdots, p_k)$. On observing that, for $k = 2$, the category of M^p is the maximum of cat M_1 and cat M_2 (the proof is similar to the proofs of (4.2) and theorem 22.2), it follows from theorem 7.1 that

If $M = M_1 \times M_2$ is an absolute neighborhood retract then cat $M = 2$ *if and only if M can be deformed into M^p and* cat $M_1 =$ cat $M_2 = 2$.

This is illustrated by the above example.

The lower bound in theorem 2.15 and the upper bound (20.1) coincide when the category of each component is exactly 2. Hence

COROLLARY 20.5. *If $M = M_1 \times \cdots \times M_k$ is an essential absolute neighborhood retract and if* cat $M_i = 2$ *for $i = 1, \cdots, k$ then* cat $M = k + 1$.

Illustrative of this corollary, the product of k spheres $S_{n_1}, \cdots, S_{n_k}$ of various dimensions has category $k + 1$. In particular the category of a k-dimensional torus is $k + 1$.

Anent the question raised above of the existence of an *essential* M with category $< 1 + \sum_{i=1}^{k}$ (cat $M_i - 1$), the method of Eilenberg [19] yields the following characterization:

If the absolute neighborhood retract $M = M_1 \times \cdots \times M_k$ is essential and if $\{B_1, B_2, \cdots, B_k\}$ is a covering of M by closed subsets, where $\text{cat}_M B_i \leqq \text{cat}\, M_i - 1$; $i = 1, \cdots, k$, *(so that* cat $M < \sum_{i=1}^{k}$ cat $M_i - (k - 1)$), *then for some i,* $\pi_i \mid B_i$ maps B_i *essentially on* M_i. (π_i denotes the projection of M into M_i.)

Suppose that for each i, $\pi_i \mid B_i$ is essential on M_i so that $\pi_i \mid B_i$ is homotopic to a mapping $\phi_i' \in M_i^{B_i}$ such that $\phi_i'(B_i)$ is a proper subset of M_i. By (20.2) there is an extension $\phi_i \in M_i^M$ of ϕ_i' which is homotopic to π_i. Hence the identity mapping $(\pi_1, \cdots, \pi_k)$ of M is homotopic to $\phi = (\phi_1, \cdots, \phi_k)$. But $\phi(M)$ is a proper subset of M, which is impossible since M is essential. Hence, for some i, $\pi_i \mid B_i$ is essential on M_i.

It is conceivable that this situation occurs. The natural mapping of a k-sphere onto a k-dimensional projective space ($k \geqq 2$) is an example of an essential mapping which raises the category. (That the category of a k-dimensional projective space is $k + 1$ follows from (31.1).

21. Covering spaces. If the connected space M is of class LC^0 and has an open covering which is contractible in M in dimension 1 then, to every subgroup ω of the fundamental group $\pi_1(M)$, there is defined a covering space M_ω of M [43 chapter 8]. Denote by p a fixed point of M so that the points of M_ω are classes $[\gamma]_\omega$ of continuous arcs with initial points p, γ and γ' belong to the same class $[\gamma]_\omega$ if $\gamma'\gamma^{-1}$ defines an element of the subgroup ω of $\pi_1(M)$.

THEOREM 21.1. *If M_ω is a covering space of M, where $C(M) \neq 0$, and X_ω denotes the set which lies over X then* $\text{cat}_{M_\omega} X_\omega \leqq \text{cat}_M X$.

Since a set lying over an open set of M is open in M_ω, it is sufficient to show that the set of M_ω lying over a set A, which is contractible in M, is contractible in M_ω. Let A be contractible in M and $h \in M^{A \times [0,1]}$ be a deformation of A in

M into p, so that $h(A, 0) = p$ and $h(x, 1) = x$. For any $x \epsilon A$ and $t \epsilon [0, 1]$ and continuous arc γ in M from p to x, define

$$\gamma'_{\gamma,t}(s) = \begin{cases} \gamma\left(\frac{s}{t}\right) & \text{for} \quad 0 \leqq s \leqq t, \\ h(x, 1 + t - s) & \text{for} \quad t \leqq s \leqq 1, \end{cases}$$

so that $\gamma'_{\gamma,t}$ is a continuous arc from p to $h(x, t)$. Observe that $[\gamma'_{\gamma,t}]_\omega$ depends on $[\gamma]_\omega$ and not on the particular arc $\gamma \epsilon [\gamma]_\omega$.

Define $h' \epsilon M_\omega^{A_\omega \times [0,1]}$ where A_ω is the set lying over A, by $h'([\gamma]_\omega, t) = [\gamma'_{\gamma,t}]_\omega$. Then $h'([\gamma]_\omega, 1) = [\gamma]_\omega$ and $h'(A_\omega, 0) \subset p_\omega$ where p_ω denotes the set of points lying over p. Thus h' is a deformation of A into p_ω. But, by (5.4), p_ω is contractible in M. Hence A_ω is contractible in M.

THEOREM 21.2. *If M_ω is a covering space of M where $h_nC(M) \neq 0$ and X_ω denotes the set which lies over X then $h_n \operatorname{cat}_{M_\omega} X_\omega \leqq h_n \operatorname{cat}_M X$.*

The proof is analogous to the proof of the preceding theorem. Let A be h_n-contractible in M and let A_ω and p_ω have the same significance as above. Let $f \epsilon M_\omega^P$, with $f(P) \subset A_\omega$, be a continuous n-dimensional complex. By hypothesis there is a mapping $h \epsilon M^{P \times [0,1]}$ such that $h(P, 0) = p$ and $h(x, 1) = \phi f(x)$, where ϕ denotes the mapping of M_ω downward into M, so that ϕf is a continuous n-dimensional complex. For any $x \epsilon P$ and $t \epsilon [0, 1]$ and continuous arc γ in M from p to $\phi f(x)$ define

$$\gamma'_{\gamma,t}(s) = \begin{cases} \gamma\left(\frac{s}{t}\right) & \text{for} \quad 0 \leqq s \leqq t, \\ h(x, 1 + t - s) & \text{for} \quad t \leqq s \leqq 1, \end{cases}$$

so that $\gamma'_{\gamma,t}$ is a continuous arc from p to $h(x, t)$.

Define $h' \epsilon M_\omega^{P \times [0,1]}$ by $h'(x, t) = [\gamma'_{\gamma,t}]_\omega$ where $[\gamma]_\omega = f(x)$. Then $h'(x, 1) = f(x)$ and $h'(P, 0) \subset p_\omega$. As we have seen that p_ω is contractible in M_ω it follows that f is homotopic to a constant. Hence A_ω is h_n-contractible in M.

In the two preceding theorems the equality need not hold. This follows, for instance, from the example: M = torus, M_ω = its universal covering space, the Euclidean plane.

22. Identifications. We next study the effect of certain identifications on the category.

THEOREM 22.1. *Let K_1 and K_2 be (closed) disjoint homeomorphic retracts of M where M is connected of uniform class LC^0 and $C(M) \neq 0$ (or $h_nC(M) \neq 0$). Let N be the space obtained from M by an identification $f \epsilon N^M$ of the corresponding points of K_1 and K_2. Then* $\operatorname{cat} M \leqq \operatorname{cat} N$ (or $h_n \operatorname{cat} M \leqq h_n \operatorname{cat} N$).

Let $\tilde{N}$ denote the space obtained from a sequence $\{M^i\}$, $i = \cdots, -1, 0, 1, \cdots$ of copies of M by identifying the corresponding points of K_2^i and K_1^{i+1} for each i, where K_1^i and K_2^i are the sets of M^i which correspond to K_1 and K_2 respectively. It is no loss of generality to suppose that $M = M^0$, $K_1 = K_1^0$,

$K_2 = K_2^0$. Thus the mapping ϕ of $\tilde{N}$ into N defined by $\phi(x) = f(x^0)$, where x^0 is the point of M^0 which corresponds to $x \in \tilde{N}$, is an extension of the mapping f. But ϕ is a covering mapping—that is to say, for every point $y \in N$ there is a neighborhood U such that $\phi^{-1}(U) = \sum V_\alpha$, where each V_α is open in $\tilde{N}$ and $\phi \mid V_\alpha$ is a topological mapping of V_α on U. It follows [43, chapter 8] that $\tilde{N}$ is a covering space of N. Hence, by theorem 21.1 (or by theorem 21.2), cat $\tilde{N} \leqq$ cat N (or h_n cat $\tilde{N} \leqq h_n$ cat N). But since K_0 and K_1 are retracts of M, $M = M^0$ is a retract of $\tilde{N}$. Hence, by (8.37), cat $M \leqq$ cat $\tilde{N}$.

That the condition that K_1 and K_2 be retracts of M cannot be dropped can be seen from the following example: M is the 2-dimensional torus, K_1 a meridian of M, K_2 a simple closed curve disjoint to K_1 which can be contracted in M. K_1 is a retract of M and K_2 is not. The category of M is 3, as we have seen earlier, but the category of N is 2. In fact it is not difficult to construct a minimal covering $\epsilon\ C(N)$ with two sets, which are images under f of cylinders each deformable in M into K_1.

THEOREM 22.2. *If K_1 and K_2 of the preceding theorem are points and if* $1 <$ cat $M < \infty$ *(or if* $1 < h_n$ cat $M < \infty$*) then* cat M = cat N *(or* h_n cat $M = h_n$ cat N*).*

Let $\{X_i\}$ be an open minimal categorical covering of M. We may assume that $K_1 \not\subset X_1$ and $K_2 \not\subset X_2$. (From the assumptions that cat $M > 1$ follows the existence of X_1 and X_2). For if, for example, $K_1 \subset \prod_{i=1}^{\text{cat } M} X_i$, choose a closed neighborhood A_1 of K_1 such that $f(A_1)$ is a categorical neighborhood of $K = f(K_1)$ in N and replace the covering $\{X_i\}$ by the open refinement $\{X_i'\}$, where

$$\begin{cases} X_1' = X_1 - A_1, & \\ X_i' = X_i, & \text{for} \quad i > 1. \end{cases}$$

This operation can be performed simultaneously if necessary on both K_1 and K_2 so that we may assume that $K_1 \not\subset X_1$ and $K_2 \not\subset X_2$.

Let U_1 and U_2 be open neighborhoods of K_1 and K_2 respectively such that $f(U_i)$ and $f(U_2)$ are contractible in N, and $U_1 \cdot U_2 = U_1 \cdot X_1 = U_2 \cdot X_2 = 0$. Let $K = f(K_1 + K_2)$.

The covering $\sigma = \{f(X_i) - K, f(U_1 + U_2)\}$ of N is open and categorical in N. Furthermore $|\sigma| = 1 +$ cat M. But $(f(X_1) - K) \cdot (f(X_2) - K) \cdot f(U_1 + U_2) = 0$ so that the nerve of σ is not a simplex. Hence, by §6, σ is not a minimal covering $\epsilon\ C(N)$, so that cat $N \leqq$ cat M.

The above proof applies word for word to the n-dimensional category.

Using induction it follows from theorem 22.2 that the categories are unchanged by a succession of point identifications.

It would be worth while to generalize theorem 22.2, somehow, to the situations of theorem 22.1 (and both theorems if possible, to the type of identification $\underset{f}{+}$ considered by Borsuk [11]. An upper bound for cat N in this direction is the following (Cf. [4, p. 277. (4)] for a special case); probably too generous:

THEOREM 22.3 *Let M be an absolute neighborhood retract in the weak sense and K_1 and K_2 be disjoint homeomorphic retracts of M. Let N be the space obtained from M by an identification $f \in N^M$ of the corresponding points of K_1 and K_2. Then* $\text{cat } N \leqq \text{cat } M + k$, *where* $k = \text{cat}_M K_1 = \text{cat}_M K_2$.

By (14.2), $\text{cat}_M K_1 \leqq \text{cat } K_1$, (for any closed covering of $C(K_1)$ is contained in an open covering of $C_M K_1$)). By (4.3), $\text{cat } K_1 \leqq \text{cat}_M K_1$. Thus $\text{cat } K_1 = \text{cat } K_2 = \text{cat}_M K_1 = \text{cat}_M K_2 = k$. For the same reason $\text{cat}_N K = \text{cat } K = k$, where $K = f(K_1) = f(K_2)$. Since $f \mid (M - (K_1 + K_2))$ is a homeomorphism, $\text{cat}_N (N - K) \leqq \text{cat}_M (M - (K_1 + K_2)) \leqq \text{cat } M$. Hence, by (4.1), $\text{cat } N \leqq \text{cat}_N (N - K) + \text{cat}_N K \leqq \text{cat } M + k$.

23. Category and the fundamental group. A generalization, due to Hurewicz, of a theorem of Borsuk [14] states that for the fundamental group to be free it is sufficient that the category be $\leqq 2$. If category is replaced by 1-dimensional category the condition becomes also necessary. Precisely:

THEOREM 23.1 *The 1-dimensional (homotopy) category of an LC^1-continuum M is $= 2$ if and only if the fundamental group $\pi_1(M)$ is free and non-vanishing.*

Suppose, first, that $h_1 \text{ cat } M = 2$. By (16.2), there is a minimal covering $\{M_1, M_2\} \in h_1C(M)$ by open, connected sets. Since M is an LC^1 continuum, $M_1 \cdot M_2$ has a finite number of components. It follows from a theorem on the fundamental group of a union [43, chapter 7; 29] that $\pi_1(M)$ is a free group with a finite number of generators; $\pi_1(M)$ does not vanish because $h_1 \text{ cat } M > 1$.

Suppose, conversely, that $\pi_1(M)$ is free and non-vanishing. Since M is an LC^1 continuum, $\pi_1(M)$ has a finite number of generators. Hence the number of generators of $\pi_1(M)$ is the coherence $r(M)$ [20, p. 175, theorem 1]. Since $\pi_1(M)$ is non-vanishing, $r(M) > 0$. Hence $M = M_1 + M_2$, where M_1 and M_2 are open and connected and $M_1 \cdot M_2$ has $1 + r(M)$ components [20, p. 172, theorem 1]. Then M_1 and M_2 must both be h_1-contractible in M, for if an element of $\pi_1(M)$ different from the identity had a representative loop in M_1, for example, then it would follow, from the above quoted theorem on the fundamental group of a union, that $\pi_1(M)$ had more than $r(M)$ generators. Hence $\{M_1, M_2\} \in h_1C(M)$ so that $h_1 \text{ cat } M \leqq 2$. But $h_1 \text{ cat } M \neq 1$ because $\pi_1(M)$ does not vanish.

It is really remarkable that $h_1 \text{ cat } M = 2$ can be characterized by means of the fundamental group $\pi_1(M)$ alone. In fact

(23.2) *The 1-dimensional homotopy category is not an invariant of the fundamental group, even over the class of complexes.*

Let T_3 denote the 3-dimensional torus obtained from the 3-dimensional cube Q_3 of Euclidean 3-space by identifying the opposite faces in the usual way, and let W denote the 2-dimensional subcomplex of T_3 which is the image, under the identification, of the boundary of Q_3. By (17.3) and corollary 20.3, $h_1 \text{ cat } T_3 = 4$, while from theorem 23.1 and (5.4) it follows that $h_1 \text{ cat } W = 3$. Nevertheless T_3 and W have isomorphic fundamental groups.

From theorem 23.1 it follows that if the LC^1 continuum M is unicoherent,

its 1-dimensional category can not be 2. For if $h_1 \operatorname{cat} M = 2$ then $\pi_1(M)$ is free; if M is unicoherent it follows that $\pi_1(M) = 0$ [20, p. 175, theorem 1] thus $h_1 \operatorname{cat} M = 1$ which is a contradiction. From this follows (cf. [14]):

(23.3) *Category and n-dimensional category are not invariants of the homology groups (arbitrary modulus) even over the class of manifolds.*

The 3-sphere S_3 and a Poincaré manifold L have the same homology groups (arbitrary modulus). It is clear that $h_1 \operatorname{cat} S_3 = 1$, $h_n \operatorname{cat} S_3 = 2$ for every $n > 1$ and $\operatorname{cat} S_3 = 2$. Since $\pi_1(L) \neq 0$, $h_1 \operatorname{cat} L > 1$. Since $b_1(L) \neq 0$, $\pi_1(L)$ is not free [39, chapter 7, §48] so that, by theorem 23.1, $h_1 \operatorname{cat} L \geqq 3$. Hence, by (17.1), $h_n \operatorname{cat} L \geqq 3$, for every n, and $\operatorname{cat} L \geqq 3$.

We can now calculate the categories of the compact, connected 2-dimensional manifolds. It is more or less obvious that $h_1 \operatorname{cat} S_2 = 1$, $h_n \operatorname{cat} S_2 = \operatorname{cat} S_2 = 2$ for $n > 1$. From theorem 23.1 and (5.4) it follows that for any other compact, connected, 2-dimensional manifold M, $h_n \operatorname{cat} M = M = 3$.

III. Homology Categories

24. Definitions. For any coefficient domain $\mathfrak{A}$ and positive integer k a mapping $\phi \in M^X$ induces a (natural) homomorphism of the k-dimensional homology group $\beta_k(X) = \beta_k(X, \mathfrak{A})$ of X (with coefficient domain $\mathfrak{A}$) into the k-dimensional homology group $\beta_k(M) = \beta_k(M, \mathfrak{A})$. (Homology groups are defined as Vietoris limit cycles [26, p. 521].) Mappings ϕ and $\psi \in M^X$ are said to be *homologous* ($\mathfrak{A}$) *in dimension* $n \geqq 1$, or *n-homologous* ($\mathfrak{A}$), if, for every k, $0 \leqq k \leqq n$, the same homomorphism of $\beta_k(X, \mathfrak{A})$ into $\beta_k(M, \mathfrak{A})$ is induced by ϕ and ψ. In other terms: ϕ and ψ are n-homologous ($\mathfrak{A}$) if every k-cycle ($\mathfrak{A}$), $0 \leqq k \leqq n$, of X is "mapped" by ϕ and ψ into homologous cycles of M [2, p. 211].

A subset X of M will be said to be H_n *deformable* ($\mathfrak{A}$) *in* M *into* Y if there is a mapping $f \in M^X$, with $f(X) \subset Y$, which is n-homologous ($\mathfrak{A}$) to the identity mapping of M^X. I shall say that X is H_n *contractible* ($\mathfrak{A}$) in M if there is a point $m \in M$ into which X can be H_n deformed ($\mathfrak{A}$) in M. In other words X is H_n-contractible ($\mathfrak{A}$) in M if every k cycle ($\mathfrak{A}$), $0 \leqq k \leqq n$, in X bounds in M.

I shall say that X is H_n *categorical* ($\mathfrak{A}$) in M if X is contained in an open set which is H_n contractible ($\mathfrak{A}$) in M. A covering of X by H_n categorical ($\mathfrak{A}$) subsets of M will be called an H_n *categorical* ($\mathfrak{A}$) *covering of* X *in* M; the collection of such coverings will be denoted by $H_nC_M(X) = H_nC_M(X, \mathfrak{A})$.

For future use let us observe that

(24.1) *If M is a complex, a subcomplex X is H_n categorical in M if and only if it is H_n contractible in M.*

The *n-dimensional homology category*, $H_n \operatorname{cat}_M X = H_n \operatorname{cat} (X, \mathfrak{A})$, of X in M is defined to be the smallest of the cardinal numbers $|\sigma|$ as σ ranges over $H_nC_M(X, \mathfrak{A})$. A covering σ of $H_nC_M(X, \mathfrak{A})$ is *minimal* if $|\sigma| = H_n \operatorname{cat}_M (X, \mathfrak{A})$.

Mappings ϕ and $\psi \in M^X$ are *homologous* ($\mathfrak{A}$) if they are homologous ($\mathfrak{A}$) in every dimension. As above, we define *H-deformable* ($\mathfrak{A}$) in M into B, *H-contractible* ($\mathfrak{A}$) in M, *H-categorical* ($\mathfrak{A}$) in M, and the *homology category*, $H \operatorname{cat}_M (X, \mathfrak{A})$ in terms of homology ($\mathfrak{A}$).

25. Relation to previously defined categories. The results of chapter I apply also to the homology category. In fact, homology in dimension n and homology are symmetric, reflexive, and transitive and satisfy (2.1), (2.2), and (2.3). When M is of class LC^0, (2.4) is satisfied. When X and M are absolute neighborhood retracts, (2.5) is satisfied. For then the Vietoris and singular homology groups are the same. Suppose ϕ_1 and $\psi_1 \in M_1^{X_1}$ are n-homologous and also ϕ_2 and $\psi_2 \in M_2^{X_2}$ are n-homologous. Let $X = X_1 \times X_2$, $M = M_1 \times M_2$, $\phi = (\phi_1, \phi_2)$, $\psi = (\psi_1, \psi_2)$. An n-cycle in X may be represented by a continuous complex $f \in X^P$ and a combinatorial n-cycle γ' on P. Let P' be a copy of P, τ a topological mapping of P' on P and γ' the cycle of P' corresponding to γ on P. As in the case of n-homotopy, $\phi f = (\phi_1\pi_1 f, \phi_2\pi_2 f)$ and $\psi f = (\psi_1\pi_1 f, \psi_2\pi_2 f)$. Since by hypothesis $\phi_1\pi_1$ and $\phi_2\pi_2$ are n-homologous to $\psi_1\pi_1$ and $\psi_2\pi_2$ respectively, there is a complex $Q \supset P + P'$ and a chain Γ on Q whose boundary is $\gamma - \gamma'$, and mappings $\alpha_1 \in M_1^Q$, $\alpha_2 \in M_2^Q$ such that $\alpha_1 | P = \phi_1\pi_1 f$, $\alpha_1 | P' = \psi_1\pi_1 f\tau$, $\alpha_2 | P = \phi_2\pi_2 f$, $\alpha_2 | P' = \psi_2\pi_2 f\tau$. Thus $\alpha = (\alpha_1, \alpha_2) \in M^Q$ such that $\alpha | P = \phi f$ and $\alpha | P' = \psi f \tau$. Hence ϕ and ψ are n-homologous, completing the proof that (2.5) is satisfied. In view of (24.1) it follows that §9 applies to homology, at least if M is a complex.

It is quite clear that $H_n \operatorname{cat}_M X \leqq H \operatorname{cat}_M X$, $H_k \operatorname{cat}_M X \leqq H_n \operatorname{cat}_M X$ if $k \leqq n$, $H_n \operatorname{cat}_M X \leqq h_n \operatorname{cat}_M X$ and $H \operatorname{cat}_M X \leqq \mathrm{h} \operatorname{cat}_M X (= \operatorname{cat}_M X)$. Also that $H_n \operatorname{cat}_M X = H \operatorname{cat}_M X$ when M is compact and n-dimensional.

26. Complete homology category. Mappings ϕ and $\psi \in M^X$ are said to be *completely homologous* if they are homologous ($\mathfrak{A}$) for every choice of coefficient domain $\mathfrak{A}$ [2, p. 211]. I shall say that ϕ and ψ are *completely n-homologous* if they are n-homologous ($\mathfrak{A}$) for every $\mathfrak{A}$. As in §24 we may define the *complete homology category* and the *complete n-homology category*. From (24.1) and [2, §4] of [27, Satz 1.3] it follows that the complete n-homology category of X in a complex M is $= H_n \operatorname{cat}_M (X, Z)$ where[20] Z is the direct sum of the cyclic groups Z_m of order $m \geqq 2$. If M is a complex without torsion then the complete n-homology category of X in M is $= H_n \operatorname{cat}_M (X, \mathfrak{R}_1)$ [2, §4, Nr. 13]; this statement may be false if M has torsion [2, §4, Nr. 13].

27. Covering spaces. The question naturally arises as to whether the analogue of theorems 21.1 and 21.2 holds for the homology categories. The following example shows that it does not:

Let M be the manifold with boundary obtained from a 2-dimensional torus by removing an open 2-cell and let X be the 1-sphere which is the boundary of M [2, VII, §1, Nr. 9]. The fundamental group of M is free with two generators, a and b, corresponding to an equator and a meridian of the original torus. Let M_ω be the covering space of M determined by the subgroup ω of $\pi_1(M)$ generated by a^2. Thus M_ω may be obtained from a torus by removing two open 2-cells

[20] I shall use, throughout, the following notation: Z_0 is the group of integers; Z_m is the cyclic group of order $m \geqq 2$; $\mathfrak{R}_1$ is the group of real numbers mod 1.

with disjoint closures. Clearly X_ω is the boundary of M_ω and consists of a pair of disjoint 1-spheres. It is not very difficult to see that $H_1 \operatorname{cat}_M X = 1$ while $H_1 \operatorname{cat}_M X = 2$, the coefficient domain $\mathfrak{A}$ chosen arbitrarily.

28. Homology categories and intersection cycles. A distinct advantage of the homology categories is the opportunity to apply the duality theorem to their calculation. In this number, the basis of this calculation is developed. Let $\mathfrak{B}$ be a locally compact separable group and let $\mathfrak{A}$ be the locally compact separable group of its characters [41, chapter V]. We require further that $\mathfrak{B}$ be a ring. We consider an orientable manifold,[21] M, of dimension n and a subcomplex A. A consequence of the duality theorem is:

If every r cycle ($\mathfrak{A}$) *in A bounds in M then every* $(n - r)$*-cycle* ($\mathfrak{B}$) *in M has a homologous cycle in* $M - A$, $(r > 0)$.

For there is a natural homomorphism θ of $\beta_r(M, \mathfrak{A})$ into $\beta_r(M \bmod A, \mathfrak{A})$, which reduces every r-cycle mod A. The hypothesis that every r-cycle in A bounds in M is equivalent, as an elementary argument shows, to the hypothesis that θ is an isomorphism of $\beta_r(M, \mathfrak{A})$ into $\beta_r(M \bmod A, \mathfrak{A})$. Let θ' denote the homomorphism, induced by θ, of the group $\beta_{n-r}(M - A, \mathfrak{B})$ of characters of $\beta_r(M \bmod A, \mathfrak{A})$ into the group $\beta_{n-r}(M, \mathfrak{B})$ of characters of $\beta_r(M, \mathfrak{A})$. A character λ of $\beta_r(M \bmod A, \mathfrak{A})$ is transformed by θ' into the character $\lambda\theta$ of $\beta_r(M, \mathfrak{A})$. It is easy to verify that θ' is the natural homomorphism of $\beta_{n-r}(M - A, \mathfrak{B})$ into $\beta_{n-r}(M, \mathfrak{B})$. Since θ is an isomorphism, there is, for every $\mu \in \beta_{n-r}(M, \mathfrak{B})$, a character λ' of a subgroup of $\beta_r(M \bmod A, \mathfrak{A})$ such that $\mu = \lambda'\theta$. But there is a character λ of $\beta_r(M \bmod A, \mathfrak{A})$ which is an extension of λ' [41, chapter 5, §31, theorem 35]. Since $\theta'\lambda = \lambda\theta = \mu$ we have shown that θ' covers the image group, i.e. $\theta'(\beta_{n-r}(M - A, \mathfrak{B})) = \beta_{n-r}(M, \mathfrak{B})$. This last statement means precisely that every $(n - r)$ cycle ($\mathfrak{B}$) in M has a homologue in $M - A$.

From this corollary of the duality theorem follows:

THEOREM 28.1 *If, on the orientable n-dimensional manifold M, there can be found k cycles* $\gamma_1, \cdots, \gamma_k(\mathfrak{B})$, *of dimensions* $\leqq n - 1$ *and* $\geqq n - r$, *such that their intersection cycle* [33, p. 171] $\gamma = \gamma_1 \cdot \gamma_2 \cdots \gamma_k$ *is not homologous to zero, then* $H_r \operatorname{cat}(M, \mathfrak{A}) \geqq k + 1$.

Suppose $H_r \operatorname{cat}(M, \mathfrak{A}) \leqq k$ so that, by (24.1), k subcomplexes $A_1, \cdots, A_k$ of M can be found, which cover M and are H_r contractible ($\mathfrak{A}$) in M. By the above argument, k homologues, $\delta_1, \delta_2, \cdots, \delta_k$, of $\gamma_1, \gamma_2, \cdots, \gamma_k$ respectively, can be found in $M - A_1, M - A_2, \cdots, M - A_k$ respectively. The intersection cycle $\delta = \delta_1 \cdot \delta_2 \cdots \delta_k$ is homologous to γ, so that by hypothesis the carrier $\hat{\delta}$ of δ is not empty; on the other hand $\hat{\delta} \subset (M - A_1)\cdot(M - A_2) \cdots (M - A_k) = 0$. This contradiction proves the theorem.

The corresponding theorem for the non-orientable manifolds has been proved by Schnirelmann [42; 39, p. 33; 40, p. 42] (in the case $r = n$) though not explicitly formulated:

[21] An n-manifold is a connected complex such that the linked complex of any r-simplex is a homology $(n - r - 1)$-sphere.

THEOREM 28.2 *If, on the n-dimensional manifold M, there can be found k cycles $\gamma_1, \gamma_2, \cdots, \gamma_k(Z_2)$ of dimensions $\leqq n - 1$ and $\geqq n - r$, such that their intersection cycle $\gamma = \gamma_1 \cdot \gamma_2 \cdots \gamma_k$ is not homologous to zero, then H_r cat $(M, Z_2) \geqq k + 1$.*

The proof is along the same lines; the group Z_2 of integers mod 2 replaces both $\mathfrak{A}$ and $\mathfrak{B}$ and mod 2 characters replace the (real numbers mod 1) characters. The extension theorem corresponding to [41, theorem 35] is trivial because subgroups of homology groups (Z_2) are closed with respect to both the topology and division.

29. Homology categories and submanifolds. In the applications of §28 to the calculation of homology categories it is sometimes sufficient to use weakened forms of theorems 28.1 and 28.2. The weaker theorem (for the case Z_2 and $r = n$) is due to Schnirelmann [42; 39, p. 32; 40, p. 40]. It is based on a sequence of manifolds with properties somewhat analogous to the categorical sequence of §5.

I shall say that a sequence $M_0, M_1, \cdots, M_k = M$ of submanifolds, of dimension $n_0, n_1, \cdots, n_k$ respectively, of an n dimensional manifold M is an S-sequence $(\mathfrak{A})$ if

$$M_0 \subset M_1 \subset \cdots \subset M_k = M \text{ and}$$

$$0 \leqq n_0 < \cdots < n_k = n \text{ and}$$

for every $i = 1, 2, \cdots, k - 1$, any $n_i - n_{i-1}$ cycle $(\mathfrak{A})$ in M_i which bounds in M bounds also in M_i.

(The last condition is somewhat analogous to the "divisor" of §4.)

The application of S-sequences to the calculation of homology category depends on the lemmas immediately below, which are consequences of the duality theorems.

A cycle γ' on a submanifold M' of a manifold M will be said to have been *extended* to a cycle γ on M if the intersection cycle $\gamma \cdot \mu'$ of γ with the fundamental cycle μ' of M' is homologous on M to γ'. This definition is for an arbitrary coefficient domain if M and M' are orientable, for coefficient domain Z_2 otherwise.

Let M be an orientable n-dimensional manifold, $\mathfrak{A}$ and $\mathfrak{B}$ as in §28.

(29.1) *If M' is an orientable m-dimensional submanifold of M, with the property that every $(m - i)$-cycle $(\mathfrak{A})$ of M' which bounds in M bounds also in M', then every i-cycle $(\mathfrak{B})$ on M' can be extended to an $(n - m + i)$-cycle $(\mathfrak{B})$ on M.*

The hypothesis that every $(m - i)$-cycle $(\mathfrak{A})$ of M' which bounds in M bounds also in M' means precisely that the natural homomorphism θ of $\beta_{m-i}(M', \mathfrak{A})$ into $\beta_{m-i}(M, \mathfrak{A})$ is an isomorphism. Let γ' be an i-cycle $(\mathfrak{B})$ on M'. According to the duality theorem, γ' is a character of $\beta_{m-i}(M', \mathfrak{A})$. Hence $\gamma'\theta^{-1}$ is a character of a subgroup of $\beta_{m-i}(M, \mathfrak{A})$. There is a character γ of $\beta_{m-i}(M, \mathfrak{A})$ which is an extension of $\gamma'\theta^{-1}$ [41, chapter 5, §31, theorem 35].

The character γ is an $(n - m + i)$ cycle $(\mathfrak{B})$ on M and $\gamma \cdot \mu'$ is homologous to γ' on M.

If we allow M and M' to be non-orientable we have in an analogous fashion (cf. §28):

(29.2) *If M' is an m-dimensional submanifold of the n-dimensional manifold M, with the property that every $(m - i)$-cycle (Z_2) of M' which bounds in M bounds also in M', then every i-cycle (Z_2) on M' can be extended to an $(n - m + i)$-cycle (Z_2) on M.*

From (29.1) and (29.2) we derive, by parallel arguments, the following two theorems of which we exhibit the proof of the first only.

THEOREM 29.3 *If there is an S-sequence $(\mathfrak{A})$, $M_0, M_1, \cdots, M_k = M$, of length $k + 1$, in the orientable n-dimensional manifold M, such that the submanifolds $M_0, M_1, \cdots, M_k$ are orientable, then H_r cat $(M, \mathfrak{A}) \geqq k + 1$ for any $r \geqq \max_{i=1,\cdots,k} \{n_i - n_{i-1}\}$.*

THEOREM 29.4 *If there is an S-sequence (Z_2), $M_0, M_1, \cdots, M_k = M$ of length $k + 1$, in the n-dimensional manifold M then H_r cat $(M, Z_2) \geqq k + 1$ for any $r \geqq \max_{i=1,\cdots,k} \{n_i - n_{i-1}\}$.*

PROOF OF THEOREM 29.3: Let $\mu_0, \mu_1, \cdots, \mu_k$ denote fundamental cycles $(\mathfrak{B})$ of $M_0, M_1, \cdots, M_k$ respectively. Applying (29.1) to the n_{i-1} dimensional cycle μ_{i-1} of M_i, $i = 1, 2, \cdots, k$, we construct an $(n - n_i + n_{i-1})$-dimensional cycle $(\mathfrak{B})$ $\overset{*}{\mu}_{i-1}$ on M which is an extension of μ_{i-1}. It is easy to prove inductively that the n_{k-i} dimensional intersection cycle $\overset{*}{\mu}_{k-1} \cdot \overset{*}{\mu}_{k-2} \cdots \overset{*}{\mu}_{k-i}$ is homologous to μ_{k-1}, and thus that the intersection cycle $\overset{*}{\mu}_{k-1} \cdot \overset{*}{\mu}_{k-2} \cdots \overset{*}{\mu}_0$ is homologous to μ_0, hence not homologous to zero. Since the dimension of each of the cycles $\overset{*}{\mu}_{k-1}, \overset{*}{\mu}_{k-2}, \cdots, \overset{*}{\mu}_0$ is $\leqq n - 1$ and $\geqq n - r$, it follows from theorem 28.1 that H_r cat $(M, \mathfrak{A}) \geqq k + 1$.

From theorem 29.3 we deduce the interesting consequence:

(29.5) *If a manifold M has dimension $\geqq 2$ and $\beta_1(M, Z_2) \neq 0$ then H cat $(M, Z_2) \geqq 3$.*

For there is a simple closed curve M_1 such that the sequence $M_0 =$ point of M_1, M_1, $M_2 = M$ is an S-sequence (Z_0).

30. Product manifolds. Consider cycles $\gamma_1^j, \cdots, \gamma_{k_j}^j (\mathfrak{B})$ of dimension $\leqq n_j - 1$ and $\geqq n_j - r_j$ on the n_j-dimensional manifold M^j $(j = 1, 2)$, where $\mathfrak{B}$ is understood to be Z_2 if the manifolds are not restricted to be orientable, such that the intersection cycles $\gamma^1 = \gamma_1^1 \cdots \gamma_{k_1}^1$ and $\gamma^2 = \gamma_1^2 \cdots \gamma_{k_2}^2$ are not homologous to zero. Then the cycles $\gamma_1^1 \times \mu^2, \cdots, \gamma_{k_1}^1 \times \mu^2, \mu^1 \times \gamma_1^2, \cdots, \mu^1 \times \gamma_{k_2}^2$, $\mu^1 \times \mu^2$ on $M^1 \times M^2$, where μ^1 and μ^2 are fundamental cycles of M^1 and M^2 respectively, have an intersection not homologous to zero. In fact this intersection is $\gamma^1 \times \gamma^2$, cf. [2, VII, §3, 5]. Thus

(30.1) *If on the orientable n_j-dimensional M_j $(j = 1, 2)$, there can be found k_j cycles $\gamma_1^j, \cdots, \gamma_{k_j}^j \cdot (\mathfrak{B})$ of dimensions $\leqq n_1 - 1$ and $\geqq n_1 - r_1$ such that their intersection cycles γ^1 and γ^2 are not homologous to zero then* H_r cat $(M^1 \times M^2, \mathfrak{A}) \geqq$

$k_1 + k_2 + 1$, *where* $r = r_1 + r_2$. *The same statement holds about not-necessarily-orientable manifolds if* $\mathfrak{A}$ *and* $\mathfrak{B}$ *are replaced by* Z_2.

Similarly one can show that

(30.2) *If there is an S-sequence* ($\mathfrak{A}$) *of orientable manifolds of length* $k_j + 1$ *on the orientable manifold* M^j ($j = 1, 2$) *then* H_r cat $(M^1 \times M^2, \mathfrak{B}) \geqq k_1 + k_2 + 1$ *where* $r = r_1 + r_2$ (cf. theorem 29.3). *The same statement holds if orientability is dropped and* $\mathfrak{A}$ *and* $\mathfrak{B}$ *are replaced by* Z_2.

Thus, in particular, the homology (Z_2) category of the product of k manifolds is $\geqq k + 1$. This is analogous to theorem 20.3.

31. The projective spaces. The most impressive application of theorem 29.4 is to the projective spaces.

(31.1) *For the n-dimensional projective space* P_n *we have* H_1 cat $(P_n, Z_2) =$ cat $P_n = n + 1$.

For there is an S-sequence (Z_2) $P_0, P_1, \cdots, P_n$ of projective spaces. It is sufficient to prove that every 1-cycle in P_i which bounds in P_{i+1} bounds in P_i ($i = 1, \cdots, n - 1$). This is true for $i = 1$ because no 1-cycle of P_1 bounds in P_2. For $i > 1$, suppose γ is a 1-cycle in P_i which bounds a 2-chain Γ in P_{i+1}. Choose a point of P_{i+1} not on P_i and not on the carrier $\hat{\Gamma}$ of Γ and project Γ from this point into P_i. The projected 2-chain has γ for its boundary.

Thus H_1 cat $(P_n, Z_2) \geqq n + 1$. But H_1 cat $(P_n, Z_2) \leqq$ cat $P_n \leqq n + 1$ by §25 and (5.4).

Let us observe that cat $P_n = n + 1$ has, as an immediate consequence, the often proved[22]

THEOREM 31.2 *If the* $(n - 1)$*-dimensional sphere* S_{n-1} *is covered with n closed sets then at least one of these sets contains a pair of antipodal points.*

Let $f \in P_n^{Q_n}$ be the standard identification of antipodal points of the n-dimensional cell Q_n and suppose that $\{A_1, A_2, \cdots, A_k\}$ is a closed covering of S_{n-1}, the boundary of Q_n, where no A_i contains any pair of antipodal points. Let a be the center of Q_n so that $\{\widehat{aA_1}, \widehat{aA_2}, \cdots, \widehat{aA_k}\}$ is a closed covering of Q_n, where $\widehat{aA_i}$ denotes the convex join of a and A_i. Since $\{f(\widehat{aA_1}), f(\widehat{aA_2}), \cdots, f(\widehat{aA_k})\}$ is a covering of P_n by closed sets which are contractible in P_n, it follows that cat $P_n \leqq k$. Hence $n + 1 \leqq k$, which proves the theorem.

32. Complexes with homology category 2. Let $\mathfrak{A}$ now denote an arbitrary coefficient domain, and let K be a connected n-dimensional complex. Suppose that H_r cat $(K, \mathfrak{A}) \leqq 2$ so that $K = K_1 + K_2$ where K_1 and K_2 are subcomplexes of K (cf. 24.1) and every i-cycle ($\mathfrak{A}$) in K_1 or in K_2 bounds in K for

---

[22] [39, p. 26, lemma 1; 8, p. 178; 16; 32; 2, XII, §4, 8, Satz XI]. Cf. also [17, 24, 23]. Another consequence is *If the n-sphere* S_n *is covered with n closed sets then at least one of these sets contains a symmetric continuum.* cf. [39, p. 26, lemma 2], where this is used as the basis of the calculation of the category of P_n.

$i = 0, 1, \cdots, r$. Under these conditions the subgroup $S_i(K)$ of $\beta_i(K)$, of these elements which can be represented as cycles $\gamma_i^1 + \gamma_i^2$ with $\gamma_i^j \subset K_j$ $(j = 1, 2)$, consists of the identity only for $i = 0, 1, \cdots, r$ [2, VII, §2, 3]. Furthermore the subgroup $N_i(K_1 \cdot K_2)$ of $\beta_i(K_1 \cdot K_2)$, consisting of cycles which bound in either K_1 or K_2, is $= \beta_i(K_1 \cdot K_2)$ for $i = 0, 1, \cdots, r$ [2, VII, §2, 4]. (The group β_0 is meant in the sense of [2, V, §1, 5].) Hence, for every $i \leqq r$, the isomorphism [2, VII, §2, 8, Satz 5] takes the form

$$\beta_i(K, \mathfrak{A}) = \beta_{i-1}(K_1 \cdot K_2, \mathfrak{A}); \qquad i \leqq r. \tag{32.1}$$

Thus we have derived a necessary condition for $\{K_1, K_2\} \in H_rC(K, \mathfrak{A})$.

From (32.1) for $i = 1$ follows

THEOREM 32.2 *If K is a connected complex and $\beta_1(K, \mathfrak{A}) \neq 0$ and is not the direct sum of groups isomorphic with $\mathfrak{A}$ then H_1* cat $(K, \mathfrak{A}) \geqq 3$.

For $\beta_0(K_1 \cdot K_2, \mathfrak{A})$ is the direct sum of t groups isomorphic with $\mathfrak{A}$, where $t + 1$ is the number of components of $K_1 \cdot K_2$.

Theorem 32.2 is the analogon of theorem 23.1. The proof of the converse breaks down, since we can no longer use the coherence, as in theorem 23.1. In fact the converse is false: Let K be the 2-dimensional torus and $\mathfrak{A} = Z_0$. It is well known that $\beta_1(K, Z_0) = Z_0 + Z_0$; nevertheless H_1 cat $(K, Z_0) = 3$, by theorem 29.4.

From theorems 29.3 and 32.2 follows:

(32.3) *If M is an orientable manifold of dimension $\geqq 2$ and p is a prime such that $\beta_1(M, Z_p) \neq 0$, then H* cat $(M, Z_p) \geqq 3$. *Thus if $\beta_1(M, Z_0) \neq 0$, the complete homology category of M is $\geqq 3$.*

For if H cat $(M, Z_p) \leqq 2$ then, by theorem 32.2, $\beta_1(M, Z_p) = Z_p + \cdots + Z_p$. There is a simple closed curve α of M which carries a non-zero element z of $\beta_1(M, Z_p)$. Since $\beta_1(\alpha, Z_p)$ is the cyclic group of order p generated by z, so that the only possible homomorphisms of $\beta_1(\alpha, Z_p)$ into $\beta_1(M, Z_p)$ other than 0 are isomorphisms, the sequence $\{M_0 = \text{point of } \alpha, M_1 = \alpha, M_2 = M\}$ is an S-sequence (Z_p). Hence by theorem 29.3, H cat $(M, Z_p) \geqq 3$. The second statement follows from §26 because $\beta_1(M, Z_0) \neq 0$ implies that $\beta_1(M, Z_p) \neq 0$ for some prime p [2, V, §4, 9, p. 235].

IV. THE STRONG CATEGORIES

33. Definitions. We have considered categories associated with each of the relations homotopy, homotopy in dimension n, homology, and homology in dimension n. These relations will be denoted by $h, h_n, H = H(\mathfrak{A}), H_n = H_n(\mathfrak{A})$. In this chapter I shall write $\mathfrak{H}$ to denote an unspecified relation ϵ the collection $\{h, h_n, H, H_n\}$.

We shall now study a category-like invariant which I have called the strong category (abbreviated cat*). This is defined by considering coverings of X by sets which are $\mathfrak{H}$-contractible (i.e. not in M but in themselves). Roughly, this amounts to confusing the set A and the space M in which it is to be $\mathfrak{H}$-contracted. Hence, it seems advisable to demand not only that the sets be

$\mathfrak{H}$-contractible in themselves but that they also possess the properties of M which were found in the preceding sections to insure a relatively complete theory. For this purpose local $\mathfrak{H}$-contractibility seems reasonable. However, the following definition is slightly more convenient: Let $\mathfrak{H}C^*(M)$ denote the collection[23] of coverings of M by neighborhood retracts of M which are $\mathfrak{H}$-contractible. (Thus when M is locally $\mathfrak{H}$-contractible so is every set of every covering of $\mathfrak{H}C^*(M)$.) An $\mathfrak{H}$-contractible neighborhood retract of M will be called a strong $\mathfrak{H}$-categorical set of M. We define the $\mathfrak{H}$-strong category,[24] $\mathfrak{H}$ cat* M, of M to be the smallest of the cardinal numbers $|\sigma|$ as σ ranges over $\mathfrak{H}C^*(M)$. (There is almost no point in considering a strong category of X in M, as, for any decent space X, it would turn out to be independent of M.) Since the points of M constitute a covering of $\mathfrak{H}C^*(M)$, the $\mathfrak{H}$-strong category is always defined. If M is an absolute neighborhood retract, its strong h-categorical sets are absolute retracts and its strong h_n-categorical sets are those neighborhood retracts whose first n-homotopy groups vanish [25, Satz 5]. A strong $\mathfrak{H}$-categorical set need not, even under the most favorable circumstances, have an $\mathfrak{H}$-contractible neighborhood. In fact the 2-cell of Alexander [1] imbedded in the obvious way in a solid torus is a strong h_1-categorical set, but no neighborhood is h_1-contractible.

34. From the definition follow immediately

(34.1) *If M_1 and M_2 are neighborhood retracts of $M_1 + M_2$ then*

$$\mathfrak{H} \text{ cat}^* (M_1 + M_2) \leqq \mathfrak{H} \text{ cat}^* M_1 + \mathfrak{H} \text{ cat}^* M_2 ;$$

and

(34.2) *h_n cat* $M \leqq$ cat* M and, if $k \leqq n$, h_k cat* $M \leqq h_n$ cat* M.* The proofs are obvious.

From a theorem of Hurewicz [25, Satz 6] follows

(34.3) *If M is an n-dimensional absolute neighborhood retract then h_n cat* $M =$ cat*M.*

Similar results are easily obtained for the strong homology categories. Note that the complete n-homology category of a complex $M = H_n$ cat* (M, Z_0), [2, V, §4, p. 228].

(34.4) *If M is an absolute neighborhood retract, and A and B are strong $\mathfrak{H}$-categorical sets of M which have only one point in common, then $A + B$ is a strong $\mathfrak{H}$-categorical set. Consequently if A' and B' are disjoint $\mathfrak{H}$-categorical sets of M then $A' + B'$ can be enlarged to a strong $\mathfrak{H}$-categorical set $A' + \alpha + B'$ by the addition of a spanning arc α.*

Since A and B are absolute neighborhood retracts so is $A + B$ [7, p. 226]. For $\mathfrak{H} = h$ the contractibility of $A + B$ is a consequence of a theorem of Aron-

[23] That the collection $\mathfrak{H}C^*(M)$ is a much more complicated invariant than $\mathfrak{H}$ cat* M is indicated by an example due to Borsuk [13] of a 2-dimensional complex, P, in 3-dimensional Euclidean space which is $\mathfrak{H}$-contractible but for which the least value of $|\sigma| > 1$, as σ ranges over $\mathfrak{H}C^*(P)$, is 3.

[24] I shall write C*(M) for $hC^*(M)$ and cat* M for h cat* (M).

zajin and Borsuk [3, p. 194]. For $\mathfrak{H} = h_n$ it follows from a theorem of Kuratowski [31, p. 277]. For $\mathfrak{H} = H$ or H_n it is more or less trivial.

REMARK: The lemma is false if M does not have suitable local properties. Example in the Cartesian plane:

$$A = \{0 \leqq x \leqq 1, y = 1\} + \{x = 0, 0 \leqq y \leqq 1\} + \sum_{n=1}^{\infty} \left\{x = \frac{1}{n}, 0 \leqq y \leqq 1\right\},$$

B the set symmetric to A with respect to the origin. Every deformation of $A + B$ leaves the origin fixed.

As a consequence of (34.4) and of the theorem of Aronzajin and Borsuk quoted above we have

(34.5) *If M is an absolute neighborhood retract and σ is a minimal covering of $\mathfrak{H}C^*(M)$, then every pair of sets of σ intersects in at least two points. For $\mathfrak{H} = h$, the intersection of any pair of sets of σ is not an absolute retract.*

35. Covering spaces. By a modification of the proof of theorem 2.19 and an extension of (34.4) can be proved

(35) *If $\tilde{M}$ is a covering space of an absolute neighborhood retract M then* $\text{cat}^* \tilde{M} \leqq \text{cat}^* M$ *and* $h_n \text{cat}^* \tilde{M} \leqq h_n \text{cat}^* M$.

36. Upper bounds for the strong category. The analogue of (5.4), namely

THEOREM 36.1 *If M is a connected n-dimensional complex then* $\text{cat}^* M \leqq n + 1$.

was proved by Borsuk [15], who observed that the theorem becomes false if M is required merely to be an n-dimensional absolute neighborhood retract.

In chapters I and II, I made successive refinements of (5.4), the ultimate refinement being theorem (18.2). I shall now show that over the class of connected complexes the analogue of theorem (18.2) does not hold.

(36.2) *There is a connected m-dimensional complex, which is simply connected and acyclic in the first k dimension, for which the strong homology (Z_0) category is $>$ the upper bound $m - k + 1$ for the category of a complex satisfying these conditions.*

In our example $m = 2$, $k = 1$ and the strong homology (Z_0) category $=$ the strong homotopy category $= 3$.

Let P_m^2 denote, for $m \geqq 2$, the 2-dimensional pseudoprojective space [2, VI, Anhang 6, 7, 8, p. 266] obtained from the 2-cell $r \leqq 1$ (written in polar coordinates (r, θ)) by identifying the m-points $(1, \theta)$, $(1, \theta + 2\pi/m), \cdots$, $(1, \theta + 2\pi(m - 1)/m))$ for each $0 \leqq \theta \leqq 2\pi/m$. The fundamental group $\pi_1(P_m^2)$ of P_m^2 is the cyclic group of order m whose generator is carried by the simple closed curve a_m which is the image, under the identification, of the boundary, $r = 1$, of the 2-cell.

Since $\beta_1(P_m^2, Z_0) = \pi_1(P_m^2)$ [43, chapter 7, §48] it follows from theorem 32.2 and (5.4) that $H_1 \text{cat}(P_m^2, Z_0) = 3$.

Let m and n be relatively prime integers and let $X = X_{m,n}$ denote the complex

obtained from P_m^2 and P_n^2 by identifying a_m and a_n and, for simplicity, write $X = P_m + P_n$. Since m and n are relatively prime it follows from a previously quoted theorem on the fundamental group of a union [43, chapter 7 §52] that X is simply connected (hence also acyclic in dimension 1). Since $\beta_2(X, Z_0) \neq 0$, X is not $H(Z_0)$-contractible, so that H cat* $(X, Z_0) \geqq 2$. Suppose $\{A, B\}$ is a covering of $HC^*(X, Z_0)$. Since H_1 cat $P_m = 3$, one of the sets $A \cdot P_m$, $B \cdot P_m$, say $A \cdot P_m$, carries a 1-cycle γ which is homologous to $q\alpha_m$, $1 \leqq q \leqq m - 1$, where α_m is the generating cycle of $\beta_1(P_m^2, Z_0)$ carried by a_m. On the other hand, since P_n is not $H(Z_0)$-contractible, $A \cdot P_n$ is a proper closed subset of P_n, so that the 1-cycle γ can not bound on $P_m + A \cdot P_n$, hence not on A. This is a contradiction since A is supposed to be $H(Z_0)$-contractible.

Thus we have shown that cat* $M \leqq m - k + 1$ is not true over the class of simply connected m-dimensional complexes M acyclic in the first k dimensions. There remains a possibility that this be true if M is further restricted to range only over pseudomanifolds (or manifolds).

In a very special case the validity of the inequality cat* $M \leqq m - k + 1$ follows from a theorem of Borsuk [12, p. 58]. The simple connectivity in this special case is a consequence of the acyclicity.

(36.3) *If the connected complex M is acyclic in dimension* 1 *and can be imbedded in* 3*-dimensional Euclidean space then* cat* $M \leqq 2$.

37. Identification of a pair of points. The conditions under which an analogue of theorem 22.1 can be proved are apparently much more restrictive.

(37) *Let X be an absolute neighborhood retract and Y the absolute neighborhood retract obtained from X by an identification, f, of a pair of distinct points, a_1 and a_2. If B is a strong $\mathfrak{H}$-categorical set of Y then $f^{-1}(B)$ is either a strong $\mathfrak{H}$-categorical set not containing a_1 or a_2 or is the union of disjoint strong $\mathfrak{H}$-categorical sets A_1 and A_2, $A_1 \ni a_1$, $A_2 \ni a_2$, according as B does or does not contain $b = f(a_1) = f(a_2)$. Hence, by* (34.4), $\mathfrak{H}$ cat* $X \leqq \mathfrak{H}$ cat* Y.

If B does not contain b then $f^{-1}(B)$ is homeomorphic with B, so that $f^{-1}(B)$ is an $\mathfrak{H}$-contractible absolute neighborhood retract contained in X.

If B contains b then, since B is $\mathfrak{H}$-contractible, there is, for any $x \in f^{-1}(B)$ an arc in B from $f(x)$ to b. Hence there is an arc in $f^{-1}(B)$ from x to either a_1 or a_2. Thus $f^{-1}(B)$ has at most two components. But if $f^{-1}(B)$ were connected it would contain an arc from a_1 to a_2 which is contrary to the assumption that B is $\mathfrak{H}$-contractible (cf. 2.4). Hence $f^{-1}(B) = A_1 + A_2$ where A_i is the component of $f^{-1}(B)$ which contains a_i $(i = 1, 2)$. Now $f|A_1$ and $f|A_2$ are homeomorphisms so that A_i is homeomorphic to $f(A_i)$, $i = 1, 2$. Hence each A_i is strong $\mathfrak{H}$-contractible. But $f(A_i)$ is a retract of B; in fact the mapping

$$\rho(y) = \begin{cases} y & \text{for } y \in f(A_i) \\ b & \text{for } y \in B - f(A_i) \end{cases}$$

is a retraction of B into $f(A_i)$. It follows [7, 6] that $f(A_i)$ and hence A_i is an absolute neighborhood retract, hence strong $\mathfrak{H}$-categorical.

38. Point Identification on an irreducibly closed complex. In direct contrast to theorem 22.2, point identification may raise the strong categories. We need the following lemma:

(38.1) *If M is an irreducibly closed n-dimensional complex* ($n \geqq 2$) *with natural domain Z_m* ($m = 0, 2, 3, \cdots$) [2, VII, §1, 3, 4, 5] *and A is a closed subset which is $H_{n-1}(Z_m)$-contractible in an open subset U of M then $M - U$ is contained in a component of $M - A$.*

Suppose $M - U$ intersects both D_1 and D_2 where $M - A = D_1 + D_2$ and D_1 and D_2 are disjoint open sets. Let V be a closed neighborhood of A which is $H_{n-1}(Z_m)$-contractible in U. Subdivide M so fine that every simplex which intersects A is contained in V. Let μ denote an irreducible cycle (Z_m) of M and let Γ be the n-chain which is zero on every n-simplex which meets $A + D_2$, and agrees with μ on every other n-simplex of M. Since the carrier $\hat{\gamma}$ of the boundary $\gamma = F(\Gamma)$ of Γ is a subset of V it follows from the construction of V that there is an n-chain Δ in U whose boundary is γ. Hence $\Gamma - \Delta$ is an n-cycle on M. Since $\hat{\Delta}\cdot(M - U) = 0$ while $\hat{\Gamma} \supset D_1\cdot(M - U) \neq 0$, the n-cycle $\Gamma - \Delta$ is $\neq 0$. On the other hand, since $\hat{\Gamma}\cdot D_2 = 0$ and since $\hat{\Delta}$ does not intersect the non-vacuous set $D_2\cdot(M - U)$ the n-cycle $\Gamma - \Delta$ is carried by a proper subset of M. But this is impossible because M is irreducibly closed. Hence $M - U$ is contained in a component of $M - A$.

THEOREM 38.2 *If X is an irreducibly closed n-dimensional complex* ($n \geqq 2$) *with natural domain Z_m* ($m = 0, 2, 3, \cdots$) *and Y is obtained from X by an identification, f, of three distinct points a_1, a_2, a_3, then H_{n-1}* cat* $(Y, Z_m) \geqq 3$.

Since $n \geqq 2$ and Y is obviously not acyclic in dimension 1, Y is not $H_{n-1}(Z_m)$-contractible. It is therefore sufficient to show that there does not exist a covering $\{Y_1, Y_2\}$ of Y by $H_{n-1}(Z_m)$-contractible absolute neighborhood retracts. Let $b = f(a_1) = f(a_2) = f(a_3)$ and $X_i = f^{-1}(Y_i)$. Let Y_1 be the set of the covering $\{Y_1, Y_2\}$ which contains b. By (37), $X_1 = X_1^1 + X_1^2 + X_1^3$, where X_1^1, X_1^2, X_1^3 are disjoint absolute neighborhood retracts and $a_j \in X_1^j$ for $j = 1, 2, 3$. Since $X_1 + X_2 = X$ the sets $X - X_1$ and $X - X_2$ are disjoint. By (38.1) $X - U$ is contained in a component of $X - X_2$ for every open neighborhood U of X_2. Hence $X - X_2$ is connected, and therefore is contained in one of the components of X_1, say X_1^1. It follows that the connected set $X_1^2 + X_1^3 + (X - X_1)$ is disjoint to $X - X_2$, so that $X_1^2 + X_1^3 + (X - X_1)$ is a subset of X_2. Thus we have constructed a connected subset of X_2 which contains both a_2 and a_3. This is a contradiction with (37).

39. Category and strong category. Theorem 38.2 enables us to show that category and strong category need not be the same, even for pseudomanifolds.

(39) *The $\mathfrak{H}$-category and the strong $\mathfrak{H}$-category are not identical over the class of pseudomanifolds. In fact for every integer $n \geqq 2$ there is an n-dimensional pseudomanifold J_n for which* cat $J_n = 2$ *but* H_{n-1} cat* $(J_n, \mathfrak{A}) =$ cat* $J_n = 3$.

The pseudomanifold J_n is obtained by identifying three distinct points of the n-sphere S_n. By theorem 22.2, cat $J_n = 2$; by theorem 38.2, H_{n-1} cat* $(J_n, \mathfrak{A}) \geqq$

3, [2, V, §4, 5 and VII, §1, 7]; by actual construction of a covering of $C^*(J_n)$, cat* $J_n \leqq 3$.

From theorem 18.2 and (36.2) we see that the example of the 2-dimensional complex $X_{m,n}$ of §36 shows that category and strong category are not identical over the class of complexes. This is an essentially weaker result than (39); however the construction of this example is interesting in this connection for its own sake, as its properties seem to depend on a different idea.

Another example of the same type is the absolute neighborhood retract B constructed by Borsuk and Mazurkiewicz [9] whose category is 2 but whose strong $H_1(\mathfrak{A})$ category is infinite. Borsuk has also constructed [10] plane curves of order 3 of arbitrarily large strong category; there is even a curve of finite order whose strong category is $> \aleph_0$ [10, p. 291].

40. Homotopy type and strong category. In contrast to (10.2), I now show that

(40) *The strong $\mathfrak{H}$-category is not an invariant of the homotopy type, even over the class of 2-dimensional pseudomanifolds.*

In fact I shall construct 2-dimensional pseudomanifolds, J and K, which belong to the same homotopy type and such that $\mathfrak{H}$ cat* $J = 3$ but $\mathfrak{H}$ cat* $K = 2$.

The pseudomanifold J will be J_2 of §39, obtained from the 2-sphere S_2 by identifying three distinct points. According to (39) $\mathfrak{H}$ cat* $J = 3$ for any of our relations $\mathfrak{H}$.

Let a_1, a_2, a_3, a_4 be four distinct points of S_2 and let K be the pseudomanifold obtained from S_2 by identifying a_1 with a_2 and a_3 with a_4. Since $\beta_1(K, \mathfrak{A})$ does not vanish $\mathfrak{H}$ cat* $K \geqq 2$. But it is easily seen by actual construction of a strong $\mathfrak{H}$-categorical covering (see figure 1) that $\mathfrak{H}$ cat* $K \leqq 2$.

In order to show that J and K belong to the same homotopy type it is convenient to imbed in 3-dimensional Cartesian space (see figure 2) as follows:

Let W_u denote the circle $(x + 2)^2 + y^2 = 4$, $z = u$ for $-2 \leqq u \leqq 2$. Let D_u denote the circle $(x + 2u)^2 + y^2 = u^2$, $z = u$ for $-2 \leqq u \leqq 0$ and the circle $(x - 2u)^2 + y^2 = u^2$, $z = u$ for $0 \leqq u \leqq 2$. Let E_u denote the circle $(x + 4(u + 1))^2 + y^2 = 4(u + 1)^2$, $z = u$ for $-2 \leqq u \leqq -1$, the point $(0, 0, u)$ for $-1 \leqq u \leqq 1$, the circle $(x - 4(u - 1))^2 + y^2 = 4(u - 1)^2$, $z = 0$ for $1 \leqq u \leqq 2$. Let T^- be a topological cylinder in the half-space $z \leqq -2$ joining the circles W_{-2} and $D_{-2} = E_{-2}$, and let T^+ be a topological cylinder in the half-space $z \geqq 2$ joining the circles W_2 and $D_2 = E_2$. It is unnecessary to further specify $T = T^- + T^+$ since the mappings f, ϕ, g_t and h_t, to be described, are identities on T.

It is clear that $J = T + \sum W_u + \sum D_u$ and $K = T + \sum W_u + \sum E_u$. (the summations extending over $-2 \leqq u \leqq 2$). In the following description the symbol $\xrightarrow{F}$ will mean that the circle on the left-hand side of the symbol is transformed into the circle on the right-hand side by a translation and an irrotational similarity.

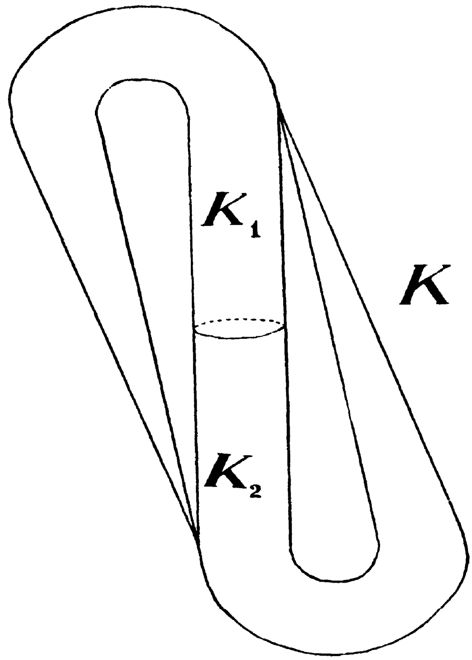


FIG. 1

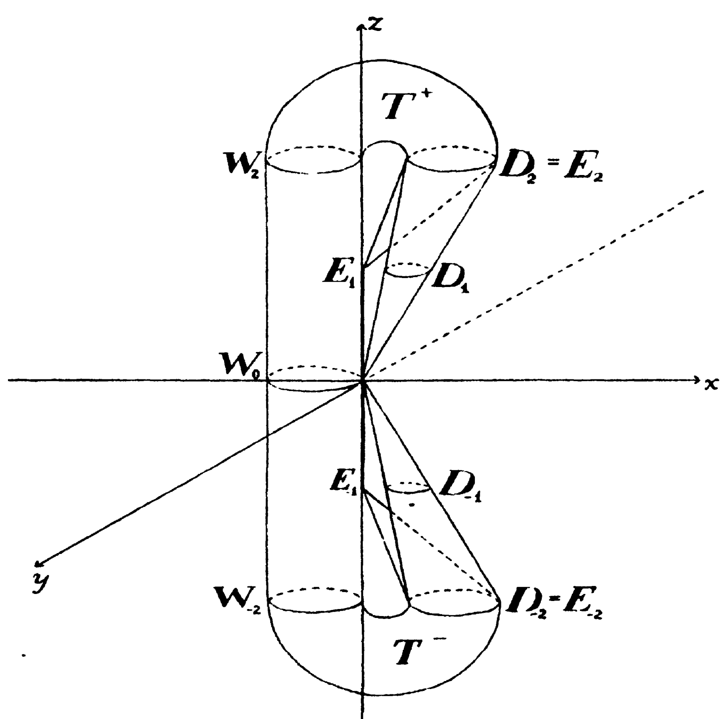


FIG. 2

The mapping $f \in K^J$ is defined by: $f \mid (T + W_u)$ is the identity, $D_u \xrightarrow{f} E_u$ for $-2 \leqq u \leqq 2$. The mapping $\phi \in J^K$ is defined by: $f \mid T$ is the identity,

$$
\begin{array}{rll}
W_u \xrightarrow{\phi} W_{2(u+1)} & \text{for } -2 \leqq u \leqq -1, \\
W_0 & \text{for } -1 \leqq u \leqq 1, \\
W_{2(u-1)} & \text{for } 1 \leqq u \leqq 2, \\
E_u \xrightarrow{\phi} D_{2(u+1)} & \text{for } -2 \leqq u \leqq -1, \\
D_{2(u-1)} & \text{for } 1 \leqq u \leqq 2.
\end{array}
$$

Mappings $g_t \in J^J$ and $h_t \in K^K$ are defined for every $t \in [0, 1]$ as follows:

$$
\begin{array}{rll}
W_u \xrightarrow{g_t} W_{(1+t)u+2t} & \text{for } -2 \leqq u \leqq -1, \\
W_{(1-t)u} & \text{for } -1 \leqq u \leqq 1, \\
W_{(1+t)u-2t} & \text{for } 1 \leqq u \leqq 2, \\
D_u \xrightarrow{g_t} D_{(1+t)u+2t} & \text{for } -2 \leqq u \leqq -1, \\
D_{(1-t)u} & \text{for } -1 \leqq u \leqq 1, \\
D_{(1+t)u-2t} & \text{for } 1 \leqq u \leqq 2.
\end{array}
$$

$$
\begin{array}{rll}
W_u \xrightarrow{h_t} W_{(1+t)u+2t} & \text{for } -2 \leqq u \leqq -1, \\
W_{(1-t)u} & \text{for } -1 \leqq u \leqq 1, \\
W_{(1+t)u-2t} & \text{for } 1 \leqq u \leqq 2, \\
E_u \xrightarrow{h_t} E_{(1+t)u+2t} & \text{for } -2 \leqq u \leqq -1, \\
E_{(1+t)u-2t} & \text{for } 1 \leqq u \leqq 2.
\end{array}
$$

We observe that $g \in J^{J \times [0,1]}$ is a homotopy between $1 \in J^J$ and $\phi f \in J^J$ and that $h \in K^{K \times [0,1]}$ is a homotopy between $1 \in K^K$ and $f\phi \in K^K$. This completes the proof that the pseudomanifolds J and K belong to the same homotopy type.

Another example of the non-dependence of the strong category on the homotopy type is the following: $X_{m,n}$ is the 2-dimensional complex of §36 (m and n always relatively prime) and Y is the pseudomanifold which is the union of 2-spheres which have exactly one point in common. By (36.2) cat* $X_{m,n} = 3$; by actual construction of a covering, cat* $Y = 2$. Since $X_{m,n}$ and Y are both simply connected and $\beta_2(X_{m,n}, Z_0)$ and $\beta_2(Y, Z_0)$ are isomorphic, $X_{m,n}$ and Y belong to the same homotopy type [27, §7]. This example is weaker than the preceding one since it shows the non-dependence of the strong $\mathfrak{H}$-category only for $\mathfrak{H} = h$ and only over the class of complexes.

41. Deformation retraction and strong category. A space and a deformation retract of it belong to the same homotopy type. It will now be shown that *the strong $\mathfrak{H}$-category is not an invariant of deformation retract over the class of*

complexes. This result which is partly weaker and partly stronger than (40) implies in particular that the analogue of (8.2) is false for strong category.

Let J be as in §40 and let $L = J + \sum_{u=-1}^{1} I_u$ where I_u denotes the convex closure of D_u, $-1 \leqq u \leqq 1$. It is trivial that J is a deformation retract of L. It remains only to show that $\mathfrak{H}$ cat* $L = 2$. But a decomposition $\epsilon\, \mathfrak{H} - C^*(L)$ is clearly $\{L_1, L_2\}$ where $L_1 = I_1 + \sum_{u=1}^{2} D_u + T + \sum_{u=-2}^{2} W_u + \sum_{u=-2}^{-1} D_u$ and $L_2 = \sum_{u=-1}^{1} I_u$.

V. Miscellany

42. Generalizations. In the preceding chapters no attempt was made to define the most general category. We have considered a class of categories which is natural and has a meaningful theory. I shall now indicate several generalizations which are natural extensions.

First: Let K be a neighborhood retract of the arcwise connected space M; a set A is categorical rel K in M if there is an open set containing A which can be deformed in M into K; a covering σ of X belongs to $C_M(X \text{ rel } K)$ if each set of σ is categorical rel K in M. The minimum, $\text{cat}_M\,(X \text{ rel } K)$, of $|\sigma|$ as σ ranges over $C_M(X \text{ rel } K)$ has many properties similar to those of $\text{cat}_M X$ to which it reduces if K is contractible in M.

We note several properties of $\text{cat}_M(X \text{ rel } K)$:

If M is an absolute neighborhood retract, a closed set is categorical rel K if and only if it can be deformed in M into K.

If X is closed and M is an absolute neighborhood retract then $\text{cat}_M\,(X \text{ rel } K) \leqq 1 + \dim (X \cdot \overline{M - K})$.

By the method of theorem 20.1 it can be shown that if $M = M_1 \times \cdots \times M_k$ is an absolute neighborhood retract and T denotes the set of points $x = (x_1, \cdots, x_k)$ of M which have at least one coordinate identical with the corresponding coordinate of a fixed point $p = (p_1, \cdots, p_n)$, then $k\, \text{cat}_M\,(M \text{ rel } T) \leqq \text{cat}\, M + k - 1$.

This last implies theorem 20.1 because if $\text{cat}\, M_i \geqq 2$ for $i = 1, \cdots, k$ and M is essential then $\text{cat}_M\,(M \text{ rel } T) \geqq 2$.

Second: Let Φ be a subset of W^M. A covering σ of X by open sets of M belongs to $C_M(X \,\|\, \Phi)$ if $\phi \mid A$ is homotopic to a constant for every $\phi \in \Phi$ and every $A \in \sigma$. As usual $\text{cat}_M\,(X \,\|\, \Phi)$ is the minimum of $|\sigma|$ as σ ranges over $C_M(X \,\|\, \Phi)$. Clearly $\text{cat}_M\,(X \,\|\, \Phi) = \text{cat}_M X$ when $W = M$ and Φ consists of the single mapping 1.

In order that a family $\Phi \in S_1^M$ be k-compatible [20, p. 158] it is necessary and sufficient that $\text{cat}_M\,(M \,\|\, \Phi) \leqq k$. Thus [20, p. 180] the multicoherence $r_k(M) \geqq n$ if and only if there exist n linearly independent mappings $f_1, f_2, \cdots, f_n \in S_1^M$ such that $\text{cat}_M\,(M \,\|\, \{f_1, \cdots, f_n\}) \leqq k$.

Furthermore [20, p. 180, theorem 1, and p. 188, theorem 1] $H_1\, \text{cat}\,(M, \mathfrak{R}_1) = \text{cat}_M\,(M \,\|\, S_1^M)$, and [20, p. 188, theorem 8] $\text{cat}_M\,(M \,\|\, S_1^M) \leqq 1 +$ maximum number of linearly independent mappings of S_1^M.

Princeton, New Jersey.

BIBLIOGRAPHY

(1) Alexander, J. W., An example of a simply connected surface bounding a region which is not simply connected. Proc. Nat. Acad. 10 (1924), 8-10.

(2) Alexandroff, P., and Hopf, H., Topologie I. Springer (1935), Berlin.

(3) Aronzajn, N., and Borsuk, K., Sur la somme et le produit combinatoire des rétractes absolus. Fund. Math. 18 (1932), 193-197.

(4) Bassi, A., Su alcuni nuovi invarianti delle varietá topologische. Annali di Math. 16 (1935), 275-297. Abstracted: On some new invariants of a manifold. Proc. Nat. Acad. 22 (1936), 698-699.

(5) Birkhoff, Garrett, Lattices and their applications. Bull. of the American Math. Soc. 44 (1938), 793-800.

(6) Borsuk, K., Sur les rétractes. Fund. Math. 17 (1931), 152-170.

(7) Borsuk, K., Über eine Klasse von lokal zusammenhängenden Räumen. Fund. Math. 19 (1932), 220-242.

(8) Borsuk, K., Drei Sätze über die n-dimensionale euklidische Sphäre. Fund. Math. 20 (1933), 177-190.

(9) Borsuk, K., and Mazurkiewicz, S., Sur les rétractes absolus indecomposables. Comptes Rend. 199 (1934), 110-112.

(10) Borsuk, K., Sur la décomposition des courbes régulières en dendrites. Fund. Math. 22 (1934), 287-291.

(11) Borsuk, K., Quelques rétractes singuliers. Fund. Math. 24 (1935), 249-258.

(12) Borsuk, K., Contribution à la topologie des polytopes. Fund. Math. 25 (1935), 51-58.

(13) Borsuk, K., Über das Phänomen der Unzerlegbarkeit in der Polyedertopologie. Comm. Math. Helvetici 8 (1935), 142-148.

(14) Borsuk, K., Über den Lusternik-Schnirelmannschen Begriff der Kategorie. Fund. Math. 26 (1936), 123-136. Abstracted: Sur la notion de la catégorie de M. L. Lusternik et Schnirelmann. Comptes Rend. 198 (1934), 1731-1732.

(15) Borsuk, K., Sur la décomposition des polyèdres n-dimensionnels en polyèdres contractiles en soi. Comp. Math. 3 (1936), 431-434.

(16) Denjoy, A., and Wolff, J., Sur la division d'une sphère en trois ensembles. Ens. Math. 32 (1933), 66-68.

(17) Eilenberg, S., Sur quelques propriétés topologiques de la surface de sphère. Fund. Math. 25 (1935), 267-272.

(18) Eilenberg, S., Sur la théorème de décomposition de la theorie de la dimension. Fund. Math. 26 (1936), 146-149.

(19) Eilenberg, S., Sur un théorème topologique de M. L. Schnirelmann. Math. Sbornik N.S. 1 (1936), 557-559.

(20) Eilenberg, S., Sur les espaces multicohérents I. Fund. Math. 27 (1936), 153-190.

(21) Eilenberg, S., Un théorème sur l'homotopie. Ann. of Math. 38 (1937), 656-661.

(22) Eilenberg, S., Sur les transformations à petites tranches. Fund. Math. 30 (1938), 92-95.

(23) Hirsch, G., Une généralisation d'un théorème de M. Borsuk concernant certaines transformations de l'Analysis Situs. Bull. Acad. Roy. Belg., 23, (1937), 219-225.

(24) Hopf, H., Freie Überdeckungen und freie Abbildungen. Fund. Math. 28, (1937) 33-57.

(25) Hurewicz, W., Topologie der Deformationen I. Proc. Akad. Amsterdam 38 (1935), 112-119.

(26) Hurewicz, W., Topologie der Deformationen II. Proc. Akad. Amsterdam 38 (1935), 521-528.

(27) Hurewicz, W., Topologie der Deformationen III. Proc. Akad. Amsterdam 39 (1936), 117-126.

(28) Hurewicz, W., Topologie der Deformationen IV. Proc. Akad. Amsterdam 39 (1936), 215-224.

(29) van Kampen, E., On the connection between the fundamental groups of some related spaces. American Journal of Math. 55 (1933), 261-267.
(30) Kuratowski, C., Topologie I. Monografje Matematyczne III (1933), Warsawa-Lwów.
(31) Kuratowski, C., Sur les espaces localement connexes et péaniens en dimension n. Fund. Math. 24 (1935), 269-287.
(32) László, R., Zu einen geometrischen Satz von K. Borsuk. Math. Fiz. Lap. 41 (1934), 36-40.
(33) Lefschetz, S., Topology. American Math. Soc. Colloquium Publications (1930), New York.
(34) Lefschetz, S., On locally connected and related sets. Ann. of Math. 35 (1934), 118-129.
(35) Lefschetz, S., On locally connected sets and retracts. Proc. Nat. Acad. 24 (1938), 392-393.
(36) Lusternik, L., and Schnirelmann, L., Sur un principe topologique en analyse. Comptes Rend. 188 (1929), 295-297.
(37) Lusternik, L., and Schnirelmann, L., Existence de trois lignés géodésiques fermées sur toute surface de genre 0. Comptes Rend. 188 (1929), 534-536, and 189 (1929), 269-271.
(38) Lusternik, L., Topologische Grundlagen der allgemeinen Eigenwerttheorie. Monatshefte für Math. u. Physik 37 (1930), 125-130.
(39) Lusternik, L., and Schnirelmann, L., Méthodes topologiques dans les problèmes variationnels. Inst. for Math. and Mechanics (1930), Mosçau. (in russian).
(40) Lusternik, L., and Schnirelmann, L., Méthodes topologiques dans les problèmes variationnels. Actualités scientifiques et industrielles (1934), Paris.
(41) Pontrjagin, L., Topological Groups. Princeton U. Press (1939) Princeton (translated from the Russian).
(42) Schnirelmann, L., Über eine neue kombinatorische Invariante. Monatshefte für Math. u. Physik 37 (1930), 131-134.
(43) Seifert, H., and Threlfall, W., Lehrbuch der Topologie. Teubner (1934), Leipzig.
(44) Wilder, R. L., On the imbedding of subsets of a metric space in Jordan continua. Fund. Math. 19 (1932), 45-64.
(a) Froloff, S., and Elsholz, L., Limite inférieure pour le nombre des valeurs critiques d'une fonction, donnée sur une variété. Math. Sbornik 42 (1935), 637-643.
(b) Gordon, J., On the minimal number of critical points of a real function defined on a manifold. Math. Sbornik N.S. 4 (1938), 105-113.
(c) Elsholz, L., Zur Theorie der Invarianten, die zur Bestimmung der unteren Grenze der Anzahl der kritischen Punkte einer stetigen Funktion, die auf einer Mannigfaltigkeit bestimmt ist, dienen können. Math. Sbornik N.S. 5 (1939), 551-558 (in russian).
(d) Elsholz, L., Die Länge einer Mannigfaltigkeit und ihre Eigenschaften. Math. Sbornik N.S. 5 (1939), 565-571 (in russian).
(e) Fox, R. H., On Homotopy and Extension of Mappings. Proc. Nat. Acad. 26 (1940), 26-28.
(f) Seifert, H., and Threlfall, W. Variationsrechnung in Grossen. Teubner (1938) Leipzig.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

ANALYTIC CURVES

BY JOACHIM WEYL

(Received August 14, 1939)

INTRODUCTION

An algebraic curve in the complex k-dimensional space $\mathfrak{R}$: $\{x_0, x_1, \cdots, x_k\}$ can be parametrically represented by setting up $x_1/x_0, x_2/x_0, \cdots, x_k/x_0$ as functions of rational character on a closed Riemann surface, the parameter being a point on the latter. From this viewpoint the algebraic curve appears as the realization of an abstract Riemann surface.

Our intention is to investigate curves $\mathfrak{C}$, defined parametrically in $\mathfrak{R}$ by setting up $x_1/x_0, x_2/x_0, \cdots, x_k/x_0$ as functions of rational character on an arbitrary Riemann surface $\mathfrak{F}$.

The methods used to accomplish this aim are based essentially upon R. Nevanlinna's theory of meromorphic functions in the generalized form developed for the investigation of meromorphic curves: H. and J. Weyl; Ann. of Math. vol. 39, pg. 516 (1938). This paper, of which the present endeavour is but an extension, will be quoted so frequently that in future it shall be referred to as "M C." In the present context the meromorphic curves appear as the special or—as we shall say—classical case where $\mathfrak{F}$ is the finite z-plane, hence $x_i = x_i(z)$ are meromorphic functions of z.

In the initial chapter we develop the first and second main theorems for the general case. In addition it contains a section concerned with the behaviour of the order of a realization of $\mathfrak{F}$ under (Kronecker-) multiplication with some other such realization, and under its projection into a lower-dimensional space. A final section discusses two specific examples; one where $\mathfrak{F}$ is an n-sheeted unbounded covering surface of the finite z-plane (Algebroid Curves); the second one where $\mathfrak{F}$ is the doubly punctured z-sphere (Ring-meromorphic Curves).

The second chapter is devoted to the defect relations (Third Main Theorem), so called because they represent the generalization of that relation holding for meromorphic functions. Their validity is shown here, besides in the classical case, only for the two specific examples mentioned above. The real addition to our theory made in this part is a modification of these relations so as to make their validity independent of the hypothesis that the exceptional points satisfy no accidental linear relations, a restriction which previously had to be imposed.

The author wishes to express in this place his gratitude and deep indebtedness to Professor H. Weyl, whose benevolent advice and frequent encouragement were essential in the completion of this paper.

I. The First and Second Main Theorems

1. Preliminary Considerations

a.) *Concerning the space* $\mathfrak{R}$:

In connection with the k-dimensional projective space $\mathfrak{R}$ we shall make use of those concepts and relations between them which were developed for the purpose of investigating the meromorphic curves in $\mathfrak{R}$. The notation employed then will be used without changes in the present treatment. To avoid repetition I shall refer the reader to the part entitled: "Distances and Means in Projective Space" of the paper[1] mentioned previously, and to H. Weyl's note on unitary metrics in projective space[2].

b.) *Concerning the surface* $\mathfrak{F}$:

Let $\mathfrak{F}$ be an arbitrary Riemann surface. Let a compact part G_0 of the surface be designated as *conductor* and fixed once and for all.

The quantities to be defined in the course of this investigation will depend on a region G whose closure is compact and which contains G_0 in its interior. Such a G shall be referred to as an admissible region. If $\mathfrak{F}$ is compact then $\mathfrak{F}$ itself is an admissible region. Primary consideration will be given to the case where both G and G_0 are connected although the actual treatment allows application to other cases as well. It will be assumed that the boundaries Γ_0 and Γ of G_0 and G, respectively, consist of a finite number of simple closed curves whose tangents vary continuously.

We now think of $\mathfrak{F}$ as a *condenser* with G_0 as the inner, charged conductor and $\bar{G} = \mathfrak{F} - G$ as the outer, grounded one. The electrostatic potential $\Phi^*(\mathfrak{p})$, defined in every point $\mathfrak{p}$ of $\mathfrak{F}$, which arises if G_0 is kept at the potential 1 and $\bar{G}$ at the potential 0, is harmonic in the intermediate dielectric $G - G_0$ and continuous on the whole surface. It follows from the principle concerning the maximum and minimum of a harmonic function that

$$0 \leqq \Phi^*(\mathfrak{p}) \leqq 1$$

throughout $\mathfrak{F}$. Consequently the density of charge

$$\rho^* = -\frac{1}{2\pi}\frac{\partial \Phi^*}{\partial n},$$

where n designates the normal directed toward the interior of the condenser, is $\geqq 0$ on Γ_0 and $\leqq 0$ on Γ. Strictly speaking we describe ρ^* in the neighborhood of any point $\mathfrak{p}$ on Γ_0 or Γ by means of a local uniformizing-parameter t of $\mathfrak{F}$ at that point. The designation $\partial/\partial n$, and later on ds—the element of arc-length of Γ_0 or Γ at $\mathfrak{p}$—, have to be understood therefore as referring to the image in

[1] M C. pp. 516–520.

[2] H. Weyl: On Unitary Metrics in Projective Space, Ann. of Math., vol. 40, no. 1, p. 141.

the t-plane. The operation $(\partial/\partial n)\, ds$ however is independent of the choice of the local parameter. This applies in particular to

$$-\frac{1}{2\pi}\frac{\partial\Phi^*}{\partial n}\, ds = d\sigma^*,$$

which is the charge of an element ds of arc-length on Γ_0 or Γ. Changing the sign in the definition of $d\sigma^*$ for an element ds on Γ, we have

$$d\sigma^* \geqq 0 \qquad \text{on } \Gamma_0 \text{ and } \Gamma.$$

The quantity

$$\int_{\Gamma_0} d\sigma^* = e$$

is the charge of the inner conductor which creates the potential Φ^*. Since the outer conductor assumes inductively the same amount of negative charge we also have

$$\int_{\Gamma} d\sigma^* = e.$$

The total energy E used to build up the charge e in the conductor G_0 is equal to the Dirichlet integral $D(\Phi^*)$ of the potential Φ^* taken over $G - G_0$ (or the whole surface $\mathfrak{F}$):

$$E = D(\Phi^*) = -\frac{1}{2\pi}\int_{\Gamma_0} \Phi^* \frac{\partial\Phi^*}{\partial n}\, ds = e.$$

Since Φ^* is not constant over the whole surface, $D(\Phi^*)$ and therefore e are actually greater than zero. The constant C with which the potential difference across $G - G_0$ has to be multiplied to obtain the charge creating it is usually referred to as the *capacity* of the condenser $\mathfrak{F}$:

$$C(\Phi_{\Gamma_0} - \Phi_{\Gamma}) = e.$$

For the particular potential Φ^* we have $C = e > 0$. Therefore we can form

$$\Phi = C^{-1}\Phi^*,$$

which is the solution of the electrostatic problem for the condenser $\mathfrak{F}$ with a normalized charge equal to unity on the conductor G_0. We obtain

$$\int_{\Gamma} d\sigma = \int_{\Gamma_0} d\sigma = 1 \tag{1.1}$$

with

$$d\sigma = -\frac{1}{2\pi}\frac{\partial\Phi}{\partial n}\, ds \qquad \text{on } \Gamma_0,$$

$$d\sigma = +\frac{1}{2\pi}\frac{\partial\Phi}{\partial n}\, ds \qquad \text{on } \Gamma.$$

If the conductor G_0 consists of several connected parts we have to think of the latter as initially being connected with each other by thin wires which will be eliminated again after the total charge has reached a distribution in equilibrium.

Let $u(\mathfrak{p})$ be a harmonic function on $\mathfrak{F}$. We apply Green's formula

$$\oint \left(u \frac{\partial \Phi}{\partial n} - \Phi \frac{\partial u}{\partial n} \right) ds = 0 \tag{1.2}$$

to the whole surface $\mathfrak{F}$. Some caution however is required because $\partial\Phi/\partial n$ has a jump

$$\lim \left\{ \left(\frac{\partial \Phi}{\partial n} \right)_- - \left(\frac{\partial \Phi}{\partial n} \right)_+ \right\} = \left[\frac{\partial \Phi}{\partial n} \right]$$

across Γ_0 as well as across Γ. The indices $-$ and $+$ refer to the values of that function at two points lying—so to speak—on opposite banks of these curves. The functions u, $\partial u/\partial n$, and Φ itself remain continuous at these junctions. Application of (1.2) results in

$$\frac{1}{2\pi} \int_\Gamma \left[\frac{\partial \Phi}{\partial n} \right] u\, ds + \frac{1}{2\pi} \int_{\Gamma_0} \left[\frac{\partial \Phi}{\partial n} \right] u\, ds = 0$$

or, since

$$\left[\frac{\partial \Phi}{\partial n} \right] ds = 2\pi\, d\sigma \quad \text{on } \Gamma, \qquad \left[\frac{\partial \Phi}{\partial n} \right] ds = -2\pi\, d\sigma \quad \text{on } \Gamma_0,$$

in the relation

$$\int_\Gamma u\, d\sigma - \int_{\Gamma_0} u\, d\sigma = 0. \tag{1.3}$$

Next consider a function F, meromorphic on $\mathfrak{F}$, that is a function which in every admissible region is of rational character. Then $u = \log | F |$ is harmonic on $\mathfrak{F}$ except for isolated singularities which it has at those points where F either possesses a pole or a zero. Before applying (1.3) we therefore cut out these isolated points by means of small circles surrounding them whose radii we let converge toward zero later on. There are only a finite number of zeros and poles of F in any admissible region; for the latter form a closed set of isolated points. The resulting equation will be

$$\int_\Gamma \log |F|\, d\sigma - \int_{\Gamma_0} \log |F|\, d\sigma = \sum_{\mathfrak{p}_0} \Phi(\mathfrak{p}_0) - \sum_{\mathfrak{p}_\infty} \Phi(\mathfrak{p}_\infty), \tag{1.4}$$

in future referred to as the fundamental equation. The sums on the right will have to be extended over all zeros $\mathfrak{p}_0$, and poles $\mathfrak{p}_\infty$ respectively of F on $\mathfrak{F}$, each counted with its proper multiplicity. But only the finite number of them which are contained in G actually contribute anything to these sums.

c.) *Concerning the curve* $\mathfrak{C}$:

An analytic curve $\mathfrak{C}$ of class $\mathfrak{F}$ in $\mathfrak{R}$ shall be defined as follows:
DEFINITION: In each point $\mathfrak{p}$ of $\mathfrak{F}$ there are given $(k + 1)$ function elements

$$x_i(\mathfrak{p}) = c_i + c_i't + \cdots, \qquad [i = 0, 1, \cdots, k], \tag{1.5}$$

t being a local uniformizing parameter of $\mathfrak{F}$ in the point $\mathfrak{p}$, with the properties

1.) $(c_0, c_1, \cdots, c_k) \neq (0, 0, \cdots, 0)$.

2.) The elements in $\mathfrak{p}'$, a point sufficiently near $\mathfrak{p}$, are obtained from the $x_i(\mathfrak{p})$ by direct analytic continuation on $\mathfrak{F}$.

3.) Effecting upon the elements (1.5) the changes:

α.) Replacement of $x_i(\mathfrak{p})$ by $\rho(\mathfrak{p})x_i(\mathfrak{p})$, where the common factor $\rho(\mathfrak{p}) = \rho_0 + \rho_1 t + \cdots$ does not vanish at $\mathfrak{p}$: $t = 0$,

β.) Replacement of t by some other local uniformizing parameter τ

$$t = g_1\tau + g_2\tau^2 + \cdots, g_1 \neq 0,$$

does not alter the point $(c_0, c_1, \cdots, c_k)$ which they define in $\mathfrak{R}$.

This point shall be called the point $\mathfrak{p}$ of the curve $\mathfrak{C}$. Furthermore we assume that $\mathfrak{C}$ does not lie in any linear subspace of $\mathfrak{R}$.

This definition makes it appear natural to look at the surface $\mathfrak{F}$ as the curve in the abstract of which $\mathfrak{C}$ is a concrete realization in the k-dimensional projective space $\mathfrak{R}$.

In order to obtain a truly geometric description of $\mathfrak{C}$ we have to do this in terms of quantities and relations which are invariant not only regarding the changes α.) and β.) but also under

γ.) an arbitrary non-singular linear transformation of the projective coordinate system:

$$y_i = \sum_j h_{ij}x_j, \qquad \det(h_{ij}) \neq 0.$$

A given linear form

$$(\alpha x) = \sum_0^k \alpha_i x_i(\mathfrak{p}), \qquad (\alpha_0, \alpha_1, \cdots, \alpha_k) \neq (0, 0, \cdots, 0),$$

will vanish to a certain finite order $h = h(\mathfrak{p}; \alpha) \geqq 0$ at $\mathfrak{p}$ if developed into a power series of the local parameter t. We shall call h the multiplicity with which $\mathfrak{C}$ cuts the plane α: $(\alpha_0, \alpha_1, \cdots, \alpha_k)$ in the point $\mathfrak{p}$. It is not affected by α.) nor β.), and, for that matter, neither by γ.) because the plane coordinates α_i of α are transformed under the latter's influence contragrediently to the point coordinates x_i.

The determinants

$$[x(\mathfrak{p})x'(\mathfrak{p}) \cdots x^{(l-1)}(\mathfrak{p})]_{i_1 \cdots i_l} = \begin{vmatrix} x_{i_1}(\mathfrak{p}) & \cdots & x_{i_l}(\mathfrak{p}) \\ \vdots & & \vdots \\ x_{i_1}^{(l-1)}(\mathfrak{p}) & \cdots & x_{i_l}^{(l-1)}(\mathfrak{p}) \end{vmatrix} \tag{1.6}$$

where the prime denotes differentiation with respect to t, assume each the factor factor ρ^l under α.) while under β.) they take on the factor

$$(dt/d\tau)^{0+1+2+\cdots+(l-1)},$$

which again amounts to the multiplication with a gauge factor different from zero at $\mathfrak{p}: \tau = 0$. It follows that the elements

$$x_{i_1 \cdots i_l}(\mathfrak{p}) = c_{i_1 \cdots i_l} + c'_{i_1 \cdots i_l} t + \cdots, \qquad [i_1 < i_2 < \cdots < i_l], \tag{1.7}$$

obtained from the determinants (1.6) by removing all factors corresponding to possible common zeros:

$$[x(\mathfrak{p}) x'(\mathfrak{p}) \cdots x^{(l-1)}(\mathfrak{p})]_{i_1 \cdots i_l} = t^{d_l} x_{i_1 \cdots i_l}(\mathfrak{p}),$$

possess properties 1.), 2.), and 3.); consequently they define a realization $\mathfrak{C}_l$ of $\mathfrak{F}$ in a space $\mathfrak{R}_l$ of $k_l = \binom{k+1}{l} - 1$ dimensions. $\mathfrak{C}_l$ is the curve $\mathfrak{C}$ defined as the locus of its generating $(l-1)$-spreads. The multiplicities $d_l = d_l(\mathfrak{p})$ are not altered by any of the changes $\alpha.$), $\beta.$), and $\gamma.$). A detailed local investigation of the manifolds $\mathfrak{C}_l$ can now be carried through along the lines followed in M C, pp. 521–522.

2. The First Main Theorem

Let $(\alpha x) = 0$ and $(\beta x) = 0$ be two planes in $\mathfrak{R}$. Into these linear forms substitute the function $x_i(\mathfrak{p})$ defining a realization $\mathfrak{C}$ of $\mathfrak{F}$ in $\mathfrak{R}$. Then the function

$$F = (\alpha x(\mathfrak{p}))/(\beta x(\mathfrak{p}))$$

will be meromorphic on $\mathfrak{F}$. Hence we can apply to it the fundamental equation (1.4). We introduce the notation

$$N(G; \alpha) = \sum \Phi(\mathfrak{p}_0) = \sum h(\mathfrak{p}; \alpha)\Phi(\mathfrak{p}),$$

where the first sum is to be extended over all zeros $\mathfrak{p}_0$ of F, each counted with its proper multiplicity; the second one however over all points $\mathfrak{p}$ of the Riemann surface $\mathfrak{F}$. Furthermore we write

$$\int_\Gamma \log \| \alpha x \|^{-1} d\sigma = m(G; \alpha),$$

$$\int_{\Gamma_0} \log \| \alpha x \|^{-1} d\sigma = m^0(G; \alpha).$$

Then (1.4) states that

$$T(G) = N(G; \alpha) + m(G; \alpha) - m^0(G; \alpha) \tag{2.1}$$

is independent of the particular plane α. This is the first main theorem for analytic curves[3]. It is to be remembered that

$$\Phi \geqq 0 \text{ on } \mathfrak{F}$$

$$d\sigma \geqq 0 \text{ on } \Gamma \text{ and } \Gamma_0 .$$

---

[3] A. Dinghas: Bemerkungen zur Ahlfors'schen Methode in der Theorie der meromorphen Funktionen, Comp. Math. 5. pp. 107–118, 1937. Dinghas generalized R. Nevanlinna's

From the latter it follows that the compensating terms $m(G; \alpha)$ and $m^0(G; \alpha)$ are $\geqq 0$. In the classical case of a meromorphic curve $m^0(G; \alpha)$ proved to be independent of $G: |z| < r$. In the general case this is no longer true, for the distribution of charge $d\sigma$ on Γ_0 will in general depend on G. m^0 may not even be bounded as a function of G. It will however be bounded when no intersections of $\mathfrak{C}$ with the plane α lie on Γ_0, and under these circumstances this is a consequence of its continuity as a function of $\alpha_0, \alpha_1, \cdots, \alpha_k$. In any case the normalization (1.1) of the charge on G_0 brings it about that the average over all planes α is

$$\mathfrak{M}_\alpha m^0(G; \alpha) = \mathfrak{M}_\alpha m(G; \alpha) = \text{const.}$$

independently of G.

The part $N(G; \alpha)$ is invariant under a transformation γ.) while the compensating terms take on under its influence an additive term lying between the bounds $\pm \log K$ independently of G, K^2 being the quotient of the maximum and minimum of the Hermitian form $\sum_0^k | \sum_0^k h_{ij} x_j |^2$ under the restriction $\sum_0^k | x_i |^2 = 1$. Let us call two functions $T_1(G)$ and $T_2(G)$ equivalent: $T_1(G) \sim T_2(G)$ if $| T_1(G) - T_2(G) |$ remains below a fixed bound for all admissible regions G. In the sense of this equivalence $T(G)$ is independent of the projective coordinate system. Thus it seems natural to say of two realizations of the curve $\mathfrak{F}$ (which are set in correspondence to each other by their co-parametrization through $\mathfrak{p}$) that they are of the same order if their T-functions are equivalent.

Next we apply the two averaging processes $\mathfrak{M}_\alpha$ and $\mathfrak{M}_\alpha^a$ [4] to the relation (2.1). Its left side, being independent of α, will not be affected at all. We shall formulate the resulting relations at once for all curves $\mathfrak{C}_l$ ($\mathfrak{C}_1 \equiv \mathfrak{C}$). Denoting by the subscript l that the quantities, thus marked, are to be formed in $\mathfrak{R}_l$ by means of the functions (1.7) as their prototypes were formed in $\mathfrak{R}$ by means of the functions (1.5), we introduce the conventions

$$\mathfrak{M}_\alpha N_l(G; \alpha) = N_l(G),$$

$$\mathfrak{M}_\alpha^a N_l(G; \alpha) = N_l^*(G; a),$$

$$\int_\Gamma \log [\![ax]\!]_l^{-1} \, d\sigma = m_l^*(G; a),$$

and note the relation

$$\mathfrak{M}_\alpha^a m_l(G; \alpha) - \mathfrak{M}_\alpha^a m_l^0(G; \alpha) = m_l^*(G; a) - m_l^{0*}(G; a).$$

characteristic $T(r)$ of a meromorphic function by weighting, for its computation, the points inside the circle $| z | < r$ with the values of a twice continuously differentiable, real-valued, but otherwise arbitrary function $\lambda(z)$, defined in every point of that circle. We preferred to generalize in the other direction, using reasonably arbitrary regions but weighting their points by the values of certain functions which are harmonic in their interior, thus avoiding the occurrence of the latter's Laplacian when applying Green's formula.

[4] M C. pp. 518–520.

Then the first main theorem appears in the form of the following three parallel expressions

$$
\begin{aligned}
T_l(G) &= N_l(G; \alpha) + m_l(G; \alpha) - m_l^0(G; \alpha), \\
T_l(G) &= N_l(G), \qquad\qquad (2.2)\\
T_l(G) &= N_l^*(G; a) + m_l^*(G; a) - m_l^{0*}(G; a).
\end{aligned}
$$

In what respects our argument so far, as well as the proofs of future results, will have to be modified in order to apply also to curves $\mathfrak{C}$ which are contained in a linear subspace of $\mathfrak{R}$ is a matter which has been discussed in the appendix of the previous paper[5]. This is of importance because the assumption that $\mathfrak{C}$ itself lies in no linear subspace of $\mathfrak{R}$ does not imply a similar behavior on the part of the curves $\mathfrak{C}_l$ with respect to the spaces $\mathfrak{R}_l$.

Let there be given two permissible regions G_1 and $G_2 : G_1 \subset G_2$, and let Φ_1^* and Φ_2^* be the corresponding potentials. Then $\Phi_1^* \leqq \Phi_2^*$ because the difference $\Phi_2^* - \Phi_1^*$ is

firstly $\geqq 0$ in the complimentary region $\bar{G}_1$,

secondly $= 0$ in G_0,

hence thirdly $\geqq 0$ on the boundary of the intermediate region $G_1 - G_0$, in the interior of which it is a regular harmonic function.

Therefore the same inequality holds throughout $G_1 - G_0$. Consequently we have in each point $\mathfrak{p}$

$$C_1\Phi_1(\mathfrak{p}) \leqq C_2\Phi_2(\mathfrak{p})$$

hence

$$C_1N(G_1\,; \alpha) \leqq C_2N(G_2\,; \alpha) \qquad (2.3)$$

or

$$\begin{vmatrix} N(G_1; \alpha) & N(G_2; \alpha) \\ C_1^{-1} & C_2^{-1} \end{vmatrix} \leqq 0.$$

From $\Phi_1^* \leqq \Phi_2^*$ it follows moreover that Φ_1^* decreases more rapidly in passing from a point on Γ_0 (where both of them are equal to unity) towards the interior of the intermediate region. Thus

$$-\frac{\partial \Phi_1^*}{\partial n} \geqq -\frac{\partial \Phi_2^*}{\partial n} \qquad \text{and} \qquad d\sigma_1^* \geqq d\sigma_2^*$$

on Γ_0, therefore $e_1 \geqq e_2$ or $C_1 \geqq C_2$. This last inequality together with (2.3) yields

$$N(G_1\,; \alpha) \leqq N(G_2\,; \alpha)$$

whence follow the same inequalities for $T(G)$ by averaging over all planes α.

[5] M C. p. 537.

$$(2.4)\qquad \begin{vmatrix} T(G_1) & T(G_2) \\ C_1^{-1} & C_2^{-1} \end{vmatrix} \leqq 0 \qquad \text{and a forteriori}$$

$$T(G_1) \leqq T(G_2).$$

In the classical case, where $\mathfrak{F}$ is the finite z-plane and G_0 and G are circles of radii r_0 and $r > r_0$ respectively around the origin, $T(G) \equiv T(r)$ was found to be an increasing convex function of $\log r$. (A function of regular type.) The generalization of the first part of this description is contained in the second one of the inequalities (2.4). The convexity with respect to $\log r$ leads in the classical case to the more stringent inequality

$$\begin{vmatrix} T(G_1) & T(G_2) & T(G_3) \\ C_1^{-1} & C_2^{-1} & C_3^{-1} \\ 1 & 1 & 1 \end{vmatrix} \leqq 0$$

for three circular regions $G_1 \subset G_2 \subset G_3$. An investigation of its validity in the general case would meet with considerable difficulty.

Complete analogy between the classical and the general case will be obtained if we exhaust $\mathfrak{F}$ by means of a sequence of admissible regions G_r. This means that r is a real parameter and that the sequence $\{G_r\}$ has the following properties:

1.) $G_r \subset G_{r'}$ if $r' > r$.

2.) If $\mathfrak{p}$ be a point of $\mathfrak{F}$ then there exists an r such that $\mathfrak{p}$ is in $G_{r'}$ for all $r' > r$.

Now we define $T(r) \equiv T(G_r)$ as the order of the curve $\mathfrak{C}$. In general $T(r)$ depends on the way we exhaust $\mathfrak{F}$ by regions G_r. This applies in particular to the classical case where it might have appeared as though one studied the curve in a specific parametrization represented by the parameter z. But for the fundamental equation (1.4) the choice of parametrization proves of no importance since this relation is invariant under one-to-one conformal transformations of the underlying Riemann surface $\mathfrak{F}$.

If $\mathfrak{F}$ is closed, and therefore compact, the curve $\mathfrak{C}$ is algebraic. No exhaustion is necessary, the outer conductor can be dispensed with, Φ^* is identically equal to unity, and the first main theorem states that the number of intersections of a plane α with the curve $\mathfrak{C}$ is independent of α.

In (1.4) it is permissible to let the inner conductor G_0 shrink to a point. The resulting form of (1.4) was generally used in its stead by previous authors on this subject.

3. Products and Projections

The simple results derived in the following sections are new also for the classical case. To prove them at once for the case of analytic curves brings about no additional difficulties. They will strengthen us in the belief that the order $T(G)$ which we introduced is really a natural concept and has all the properties which we expect from an "order."

a.) *Products*:

The result obtained in this section is essentially a generalization of the fact that two algebraic curves of orders m and n respectively intersect in mn points.

Consider two realizations $\mathfrak{C}$ and $\mathfrak{D}$ of the same abstract curve $\mathfrak{F}$, one of which lies in a k-space, the other one in an h-space.

$$\mathfrak{C}\colon\ x_0 : x_1 : x_2 : \cdots : x_k = x_0(\mathfrak{p}) : x_1(\mathfrak{p}) : \cdots : x_k(\mathfrak{p}),$$

$$\mathfrak{D}\colon\ y_0 : y_1 : \cdots : y_h = y_0(\mathfrak{p}) : y_1(\mathfrak{p}) : \cdots : y_h(\mathfrak{p}).$$

By Kronecker-multiplication we obtain a third realization

$$z_{ij}(\mathfrak{p}) = x_i(\mathfrak{p}) y_j(\mathfrak{p})$$

in a space of $(kh + k + h)$ dimensions. We shall call it the direct product $\mathfrak{C} \times \mathfrak{D}$ of the realizations $\mathfrak{C}$ and $\mathfrak{D}$.

Now let us compare the orders of these three curves with each other. We shall change our notation and for the order $T(G)$ of $\mathfrak{C}$ we shall write $T[\mathfrak{C}]$ assuming that thus denoted orders of different realizations are taken with respect to the same admissible region.

For the computation of $T[\mathfrak{C}]$, $T[\mathfrak{D}]$ and $T[\mathfrak{C} \times \mathfrak{D}]$ we choose the planes

$$\alpha\colon\ (\alpha_0, \alpha_1, \cdots, \alpha_k) \qquad \text{in the } k\text{-space},$$

$$\beta\colon\ (\beta_0, \beta_1, \cdots, \beta_h) \qquad \text{in the } h\text{-space},$$

and

$$\gamma\colon\ \gamma_{ij} = \alpha_i \beta_j \qquad \text{in the product-space}$$

respectively. Calculating each order for the same region G we readily obtain from

$$\left(\textstyle\sum_0^k \alpha_i x_i\right)\left(\textstyle\sum_0^k \beta_j x_j\right) = \textstyle\sum_{ij} \gamma_{ij} z_{ij},$$

$$\left(\textstyle\sum |x_i|^2\right)\left(\textstyle\sum |y_j|^2\right) = \textstyle\sum_{ij} |z_{ij}|^2, \qquad \left(\textstyle\sum |\alpha_i|^2\right)\left(\textstyle\sum |\beta_i|^2\right) = \textstyle\sum_{ij} |\gamma_{ij}|^2,$$

the relations

$$N[\mathfrak{C} \times \mathfrak{D}] = N[\mathfrak{C}] + N[\mathfrak{D}],$$

$$m[\mathfrak{C} \times \mathfrak{D}] = m[\mathfrak{C}] + m[\mathfrak{D}].$$

Hence

$$T[\mathfrak{C}] + T[\mathfrak{D}] = T[\mathfrak{C} \times \mathfrak{D}]. \tag{3.1}$$

The order of a direct product of realizations of an abstract curve $\mathfrak{F}$ is the sum of the orders of the factors.

A similar relation does not hold for the orders of higher rank.

b.) *Powers*:

We apply the theorem (3.1) to the n times reiterated product of the curve $\mathfrak{C}$ with itself. Let us define this realization $\mathfrak{C}^n$ of $\mathfrak{F}$ by the functions

$$x_{m_0 m_1 \cdots m_k}(\mathfrak{p}) = \left\{\frac{n!}{m_0!\, m_1! \cdots m_k!}\right\}^{\frac{1}{2}} x_0^{m_0}(\mathfrak{p}) \cdots x_k^{m_k}(\mathfrak{p}), \tag{3.2}$$

where $m_0, m_1, \cdots, m_k$ run over all non-negative integers whose sum $\sum_0^k m_i$ equals n. (The advantage of this choice over $x_{m_0 \cdots m_k}(\mathfrak{p}) = x_0^{m_0}(\mathfrak{p}) \cdots x_k^{m_k}(\mathfrak{p})$ will become apparent as the discussion proceeds.) We calculate the order of $\mathfrak{C}$ for a plane $(\alpha x) = 0$. On the other hand we remark that $[(\alpha x)]^n = [\sum_0^k \alpha_i x_i]^n$ is a linear combination of the monomials (3.2) with the coefficients

$$\alpha_{m_0 m_1 \cdots m_k} = \left\{\frac{n!}{m_0!\, m_1! \cdots m_k!}\right\}^{\frac{1}{2}} \alpha_0^{m_0} \alpha_1^{m_1} \cdots \alpha_k^{m_k}.$$

Making use of this for the computation of $T[\mathfrak{C}^n]$ we obtain at once

$$N[\mathfrak{C}^n] = nN[\mathfrak{C}]$$

and

$$m[\mathfrak{C}^n] = nm[\mathfrak{C}]$$

since

$$\sum |\alpha_{m_0 m_1 \cdots m_k}|^2 = \left\{\sum_0^k |\alpha_i|^2\right\}^n,$$

$$\sum |x_{m_0 m_1 \cdots m_k}|^2 = \left\{\sum_0^k |x_i|^2\right\}^n.$$

The result

$$T[\mathfrak{C}^n] = nT[\mathfrak{C}]$$

permits the following interpretation: Every plane

$$\sum \alpha_{m_0 m_1 \cdots m_k} x_{m_0 m_1 \cdots m_k} = 0$$

in the product-space can at the same time be thought of as an algebraic surface of order n in the space $\mathfrak{R}$ defined by

$$\sum \alpha_{m_0 m_1 \cdots m_k} x_0^{m_0} x_1^{m_1} \cdots x_k^{m_k} = 0.$$

In either case the sum is to be extended over all sequences of non-negative integers $m_0, m_1, \cdots, m_k$ whose sum is n. Viewed in this light $T[\mathfrak{C}^n]$ is seen to describe the behaviour of $\mathfrak{C}$ with respect to an algebraic surface of order n in the same sense that $T[\mathfrak{C}]$ describes its behaviour with respect to a plane. We found that: *The order of $\mathfrak{C}$ referred to an algebraic suface of order n is equal to n times the order of $\mathfrak{C}$ (referred to a plane).* This implies for instance that the average density of intersections of $\mathfrak{C}$ with an algebraic surface of order n is n times the average density of its intersections with a plane. (Without the numerical factor $\{n!/m_0!m_1! \cdots m_k!\}^{\frac{1}{2}}$ the final result would have appeared less sharply in the form of an equivalence $T[\mathfrak{C}^n] \sim nT[\mathfrak{C}]$.)

c.) *Projections*:

The projection from a given $(k - n - 1)$-dimensional linear subspace $\mathfrak{R}^*$ of $\mathfrak{R}$ as center into a linear subspace $\mathfrak{R}'$ of n dimensions which does not intersect $\mathfrak{R}^*$ can be described in a suitably chosen projective coordinate system by the

passage from the point $(x_0, x_1, \cdots, x_n, x_{n+1}, \cdots, x_k)$ to the point $(x_0, x_1, \cdots, x_n, 0, \cdots, 0)$.

Under the influence of such a projection the curve $\mathfrak{C}$ passes into a curve $\mathfrak{C}'$ in $\mathfrak{R}'$. If a point $\mathfrak{p}$ of the original curve lies on $\mathfrak{R}^*$ then the elements

$$x_0(t),\ x_1(t),\ \cdots,\ x_n(t) \tag{3.3}$$

will all contain a common factor $t^b (b = b(\mathfrak{p}))$; only after its elimination do the power series (3.3) define the corresponding element of the projected curve. We consider a plane α of the special form

$$\alpha_0 x_0 + \alpha_1 x_1 + \cdots + \alpha_n x_n.$$

Then it follows that

$$h'(\mathfrak{p}; \alpha) = h(\mathfrak{p}; \alpha) - b(\mathfrak{p}),$$

where h refers to $\mathfrak{C}$ and h' to $\mathfrak{C}'$. Consequently

$$h'(\mathfrak{p}; \alpha) \leqq h(\mathfrak{p}; \alpha) \qquad \text{and also}$$

$$N'(G; \alpha) \leqq N(G; \alpha).$$

The same inequality for the m-part follows directly from the fact that

$$\frac{\sum_0^n |\alpha_j|^2 \sum_0^n |x_j|^2}{|\sum_0^n \alpha_j x_j|^2} = \left(\frac{1}{\|\alpha x\|'}\right)^2 \leqq \left(\frac{1}{\|\alpha x\|}\right)^2 = \frac{\sum_0^n |\alpha_j|^2 \sum_0^k |x_i|^2}{|\sum_0^n \alpha_j x_j|^2}$$

yielding

$$T' \leqq T + (m^0 - m^{0\prime}).$$

With $X^2 = \sum_0^k |x_i|^2$, $X'^2 = \sum_0^n |x_j|^2$, we have

$$m^0 - m^{0\prime} = \int_{\Gamma_0} \log (X:X')\, d\sigma$$

and the final result now appears in the form

$$T\{\mathfrak{R}'\} \leqq T\{\mathfrak{R}\} + \int_{\Gamma_0} \log (X:X')\, d\sigma \tag{3.4}$$

which is independent of the particular choice of the plane α. The notation used explains itself.

The additional term on the right side is certainly bounded as a function of G if none of the points which $\mathfrak{C}$ has in common with $\mathfrak{R}^*$ lie on Γ_0. In the classical case the independence of m^0 from G prevents this difficulty from arising altogether. The contents of the inequality (3.4) can be expressed with due qualifications in the form: *Projection cannot essentially increase the order of an analytic curve.*

The relation (3.4) has its counter-part for higher l. For our projection $\mathfrak{C}'$ we can form the l-curves $(\mathfrak{C}')_l$. They will lie in spaces $\mathfrak{R}_l'$ of dimensionality

$\binom{n+1}{l} - 1$, $(l \leqq n)$. We may think of these spaces $\mathfrak{R}'_l$ as linear subspaces of the corresponding $\mathfrak{R}_l$ respectively. The projection of $\mathfrak{R}$ from $\mathfrak{R}^*$ induces for higher l the projection of $\mathfrak{R}_l$ from $\mathfrak{R}^*_l$ which is effected through replacing in the determinants

$$[xx' \cdots x^{(l-1)}]_{i_1 \cdots i_l}$$

all $(k + 1)$-uples $x, x', \cdots$ by the corresponding ones with zeros after the $(n + 1)$st place. This projection throws $\mathfrak{C}_l$ into $\mathfrak{R}'_l$ where it is seen to coincide with $(\mathfrak{C}')_l$: The l-curves $(l \leqq n)$ of the projection are projections of the corresponding $\mathfrak{C}_l$.

Hence writing

$$X_l^2 = \sum |x_{i_1 \cdots i_l}|^2 \qquad [i_1 < i_2 < \cdots < i_l]$$

and putting $X_l'^2$ equal to what arises from X_l^2 under the projection of $\mathfrak{R}_l$ from $\mathfrak{R}^*_l$, induced by the production of $\mathfrak{R}$ from $\mathfrak{R}^*$, we obtain also

$$T\{\mathfrak{R}'_l\} \leqq T\{\mathfrak{R}_l\} + \int_{\Gamma_0} \log (X_l : X'_l)\, d\sigma$$

which by its very form proves to be independent not only of any particular plane but also—in the sense of equivalence—of the particular coordinate system employed to derive it.

We are led to a kind of inverse of the above relations when we consider the following situation: By projection from the centers $\mathfrak{R}^*$ and $\mathfrak{R}^{**}$ the total space $\mathfrak{R}$ be mapped into $\mathfrak{R}'$ and $\mathfrak{R}''$ respectively. By $\mathfrak{R}' \vee \mathfrak{R}''$ and $\mathfrak{R}' \wedge \mathfrak{R}''$ let us denote the spaces into which $\mathfrak{R}$ turns by projection from the centers $\mathfrak{R}^* \cap \mathfrak{R}^{**}$ (intersection of $\mathfrak{R}^*$ and $\mathfrak{R}^{**}$) and $\mathfrak{R}^* \cup \mathfrak{R}^{**}$ (sum, union of $\mathfrak{R}^*$ and $\mathfrak{R}^{**}$) respectively. We assume in particular that $\mathfrak{R}^* \cap \mathfrak{R}^{**}$ is empty, then

$$\mathfrak{R}' \vee \mathfrak{R}'' = \mathfrak{R} \qquad \text{and we have}$$

$$\dim (\mathfrak{R}' \wedge \mathfrak{R}'') + \dim (\mathfrak{R}' \vee \mathfrak{R}'') = \dim (\mathfrak{R}') + \dim (\mathfrak{R}'').$$

We shall show that the relation

$$T\{\mathfrak{R}' \wedge \mathfrak{R}''\} + T\{\mathfrak{R}' \vee \mathfrak{R}''\} \leqq T\{\mathfrak{R}'\} + T\{\mathfrak{R}''\},$$

holding with proper qualifications, connects the corresponding orders. Two projections from the centers $\mathfrak{R}^*$ and $\mathfrak{R}^{**}$ are called complimentary if not only

$$\mathfrak{R}^* \cap \mathfrak{R}^{**} = 0$$

but also
$$\dim (\mathfrak{R}^* \cup \mathfrak{R}^{**}) = \dim (\mathfrak{R}) - 1.$$

Then we have, since $\mathfrak{R}' \vee \mathfrak{R}'' = \mathfrak{R}$, the relation

$$\dim (\mathfrak{R}) = \dim (\mathfrak{R}') + \dim (\mathfrak{R}'')$$

and
$$T\{\mathfrak{R}\} \leqq T\{\mathfrak{R}'\} + T\{\mathfrak{R}''\}$$

The order of an analytic curve does not essentially exceed the sum of orders obtained when adding the orders of any two complimentary ones of its projections.

To prove this choose the coordinates so that the vanishing of

$$x_0, x_1, \cdots, x_r \tag{3.5}$$

defines $\mathfrak{R}^* \cup \mathfrak{R}^{**}$, whereas

$$x_0, x_1, \cdots, x_r, \qquad y_{r+1}, \cdots, y_s, \tag{3.6}$$

$$x_0, x_1, \cdots, x_r, \qquad z_{r+1}, \cdots, z_t, \tag{3.7}$$

are the coordinates associated in the same manner with $\mathfrak{R}^*$ and $\mathfrak{R}^{**}$ respectively. The sequence of the x, y, and z together is under these circumstances a full coordinate system of $\mathfrak{R}$. The quantities (3.5), (3.6), and (3.7) are the coordinates in $\mathfrak{R}^0 = \mathfrak{R}' \wedge \mathfrak{R}''$, $\mathfrak{R}'$, and $\mathfrak{R}''$ respectively. We choose a plane

$$\alpha_0 x_0 + \alpha_1 x_1 + \cdots + \alpha_r x_r = 0.$$

Then in the same manner as before

$$h^0 = h(\mathfrak{p}; \alpha) - b^0(\mathfrak{p}), \qquad h' = h(\mathfrak{p}; \alpha) - b'(\mathfrak{p}), \qquad h'' = h(\mathfrak{p}; \alpha) - b''(\mathfrak{p}).$$

Obviously both b' and b'' are $\leqq b^0$, moreover, since there is no point where all coordinates (3.6) and (3.7) are zero, one of the two numbers b' or b'' is necessarily zero, e.g. $b'' = 0$. In combining this with $b' \leqq b^0$ one gets the relation $b' + b'' \leqq b^0$ holding for both alternatives, consequently

$$h' + h'' \geqq h + h^0.$$

This leads to

$$N' + N'' \geqq N + N^0.$$

The corresponding relation for the m-part

$$m' + m'' \geqq m + m^0$$

follow at once from the inequality

$$X'^2 + X''^2 \geqq X^2 + X^{02},$$

where the quantities X^0, X', and X'' are the projections of X from the spaces $\mathfrak{R}^* \cup \mathfrak{R}^{**}$, $\mathfrak{R}^*$, and $\mathfrak{R}^{**}$ respectively. The exact form of the desired result is therefore given in the formula

$$T\{\mathfrak{R}'\} + T\{\mathfrak{R}''\} \geqq T\{\mathfrak{R}\} + T\{\mathfrak{R}' \wedge \mathfrak{R}''\} + \gamma^0$$

with

$$\gamma^0 = \int_{\Gamma_0} \log (X'X'' : XX^0)\, d\sigma,$$

which is independent of the particular plane used to derive it.

4. The Second Main Theorem

We first derive the second main theorem under the hypothesis that in addition to our curve $\mathfrak{C}:x_i = x_i(\mathfrak{p})$ we have a meromorphic function $z(\mathfrak{p})$ on $\mathfrak{F}$ at our disposal. Then $\mathfrak{F}$ may be looked upon as a covering surface of the z-plane. Another more appropriate interpretation would be to think of $z(\mathfrak{p})$ as a curve given in addition to $\mathfrak{C}$,—a realization of $\mathfrak{F}$ in a one-dimensional space. With the abbreviations

$$\alpha(xx' \cdots x^{(l-1)}) = \sum \alpha_{i_1 \cdots i_l}[xx' \cdots x^{(l-1)}]_{i_1 \cdots i_l}, \qquad [i_1 < i_2 < \cdots < i_l],$$

we form the expressions

$$F = \frac{\alpha(xx' \cdots x^{(l-2)}) \cdot \gamma(xx' \cdots x^{(l)})}{\{\beta(xx' \cdots x^{(l-1)})\}^2} \tag{4.1}$$

interpreting the prime as signifying derivation with respect to z. This is a meromorphic function on $\mathfrak{F}$. If we hold on to the convention that the prime indicates differentiation with respect to the local parameter t, then

$$F = \frac{\alpha(xx' \cdots x^{(l-2)}) \cdot \gamma(xx' \cdots x^{(l)})}{\{\beta(xx' \cdots x^{(l-1)})\}^2} \cdot \frac{dt}{dz}.$$

The expansion of dz/dt in terms of t at the point $\mathfrak{p}$ will start with a certain power t^j $(j = j(\mathfrak{p}))$. We introduce the quantity

$$\sum j(\mathfrak{p})\Phi(\mathfrak{p}) = J^z(G)$$

and the customary

$$V_l(G) = \sum \{d_{l+1}(\mathfrak{p}) - 2d_l(\mathfrak{p}) + d_{l-1}(\mathfrak{p})\}\Phi(\mathfrak{p}),$$

the sums extending over all points $\mathfrak{p}$ of $\mathfrak{F}$. The latter may be looked upon as a measure of the density of stationary $(l - 1)$-spreads of the realization $\mathfrak{C}$ over the part G of the abstract curve. Application of the fundamental equation (1.4) to the function yields in the same manner as in the classical case the second main theorem:

$$V_l(G) + \{T_{l+1}(G) - 2T_l(G) + T_{l-1}(G)\} = J^z(G) + \{{}^z\Omega_l(G) - {}^z\Omega_l^0(G)\} \tag{4.2}$$

where

$${}^z\Omega_l(G) = \int_\Gamma \log\left\{\left(\frac{X_{l+1}X_l}{X_l^2}\right)\left|\frac{dt}{dz}\right|\right\} d\sigma$$

and where ${}^z\Omega_l^0$ denotes the same integral extended over Γ_0. The expression under the integral sign is independent of both the gauge factor and the choice of the local parameter. The entire left side of (4.2) does not depend on the particular function z. The fact that the right member also has the same value for two meromophic function z and ζ is realized by applying the fundamental equation (1.4) to the function $d\zeta/dz$.

In order to obtain a formulation of the second main theorem which does not involve the auxiliary realization $z(\mathfrak{p})$ of $\mathfrak{F}$ we argue as follows: According to the theory of uniformization the surface $\mathfrak{F}$ is one of the spatial forms which the Euclidean, the spherical, or the Lobatschewskian plane may assume and is therefore endowed with a uniquely determined Riemannian line element $d\mathfrak{s}$. By means of this line element we can form the following expressions

$$\Omega_l(G) = \int_\Gamma \log\left(\frac{X_{l+1}X_{l-1}}{X_l^2}\cdot\frac{|dt|}{d\mathfrak{s}}\right)d\sigma$$

which depend neither on the gauge factor nor on the local parameter. Now consider two realizations $\mathfrak{C}$ and $\mathfrak{D}$ of $\mathfrak{F}$ and form the quotient of two expressions (4.1) for $\mathfrak{C}$ and $\mathfrak{D}$. We are not even forced to choose the same rank l in both cases. By application of the fundamental equation (1.4) to this quotient which does not depend on the local parameter we find that

$$V_l(G) + \{T_{l+1}(G) - 2T_l(G) + T_{l-1}(G)\} - \{\Omega_l(G) - \Omega_l^0(G)\} = \epsilon(G)$$

is independent of the choice of the curve $\mathfrak{C}$ as well as of l and therefore determined only by the Riemann surface $\mathfrak{F}$. From a theoretical standpoint this relation

$$V_l(G) + \{T_{l+1}(G) - 2T_l(G) + T_{l-1}(G)\} = \Omega_l(G) - \Omega_l^0(G) + \epsilon(G) \tag{4.3}$$

is the more satisfactory form of the second main theorem, although, for the purpose of our future estimates the equations (4.2) will prove the more useful of the two. If we apply (4.3) to $\mathfrak{C}$ and to the one-dimensional curve defined by z, we are led back by subtraction to the previous form.

In the algebraic case of a compact surface $\mathfrak{F}$ we shall take $G \equiv \mathfrak{F}$. Then the integrals along the boundaries vanish and we obtain the well-known Plücker formulae stating that

$$v_l + (n_{l+1} - 2n_l + n_{l-1})$$

is a constant independent of l and the curve.

5. Examples

A general discussion proceeding along the lines of the previous sections, which now would lead naturally to an estimate of first $\Omega_l(G)$ and then $\Omega_l(G)$ augmented by defect sums, meets with forbidding difficulties. We shall limit ourselves therefore to the discussion of two examples which possess the same rotational symmetry as the classical case and therefore allow description by polar coordinates. The regions G which are used to exhaust $\mathfrak{F}$ will be bounded by concentric circles. All quantities involved in our relations will become real-valued functions of the radii of those circles. This enables us to apply the methods of estimation which were employed in the classical case to derive the defect relations. For the moment let us specialize the results gained so far to fit the cases of the proposed examples.

a.) *Algebroid Curves*:

In this case $\mathfrak{F}$ is an n-sheeted covering surface over the open z-plane without relative boundaries. The conductor G_0 shall be that part of $\mathfrak{F}$ which covers the circle $|z| \leqq r_0$ of the z-plane. The parts G_r of $\mathfrak{F}$ which cover the circles $|z| < r$ make up the sequence $\{G_r\}$, used to exhaust $\mathfrak{F}$. A complete treatment of this case has been given, though from a different viewpoint, by E. Ullrich.[6]

In our case

$$\Phi(\mathfrak{p}) = \frac{1}{n}\log\frac{r}{r_0} \qquad \text{for } \mathfrak{p} \text{ in } G_0,$$

$$\Phi(\mathfrak{p}) = \frac{1}{n}\log\frac{r}{|z|} \qquad \text{for } \mathfrak{p} \text{ in } G_r \ (z \text{ being its trace in the } z\text{-plane}),$$

$$\Phi(\mathfrak{p}) = 0 \qquad \text{for } \mathfrak{p} \text{ in } \bar{G}_r,$$

and

$$d\sigma = \frac{1}{2\pi n}\, d\varphi.$$

Indicating by $\int_0^{2\pi n}$ the integration around the boundary of G, that is, writing

$$m_l(G; \alpha) = \frac{1}{2\pi n}\int_0^{2\pi n} \log \|\alpha x\|_l^{-1}\, d\varphi = m_l(r; \alpha)$$

and

$$m_l^*(G; a) = \frac{1}{2\pi n}\int_0^{2\pi n} \log [\![ax]\!]_l^{-1}\, d\varphi = m_l^*(r; a),$$

we obtain the first main theorem for algebroid curves in the form

$$T_l(r) \sim N_l(r; \alpha) + m_l(r; \alpha) \sim N_l(r) \sim N_l^*(r; a) + m_l^*(r; a)$$

because

$$m^0(r; \alpha) = m(r_0\,; \alpha),$$

formed by integrating around Γ_0, is independent of r and therefore bounded, which permits us to write the customary equivalences.

The second main theorem is best stated in the form (4.2) with z as the variable. We note immediately that ${}^z\Omega_l^0(G)$ is bounded for all G_r. Hence (4.2) appears, written as an equivalence, in the form

$$V_l(r) + \{T_{l+1}(r) - 2T_l(r) + T_{l-1}(r)\} \sim J(r) + \Omega_l(r).$$

[6] E. Ullrich: Wertverteilung und Verzweigtheit von Algebroiden, Crelle's Journal für die reine und angew. Mathematik, Bd. 167, p. 198.

For the quantity $\Omega_l(r) = {}^e\Omega_l(G_r)$, we have in keeping with the previously adopted notation,

$$\Omega_l(r) = \frac{1}{2\pi n}\int_0^{2\pi n} \log\left(\frac{X_{l+1}X_{l-1}}{X_l^2}\right) d\varphi,$$

while $J(r)$ measures the density of branchpoints of $\mathfrak{F}$ over the disc $|z| < r$ of the z-plane.

b.) *Ring-meromorphic Curves*:

As a second example we shall consider the curves $\mathfrak{C}$ defined by $(k + 1)$ functions

$$x_0 = x_0(z), \qquad x_1 = x_1(z), \cdots, \qquad x_k = x_k(z),$$

which are meromorphic on the doubly punctured z-plane. Let us assume that $\mathfrak{F}$ is the z-plane punctured at $z = 0$ and $z = \infty$. The natural choice of our region G is a ring bounded by two concentric circles of radii R and r, $R > r$, respectively. The first one excludes a neighbourhood of $z = \infty$, the second one a neighbourhood of $z = 0$. Inside the region G we have to fix the conductor G_0 which is to contain the charge creating the potential Φ. We take, as the most convenient one, another ring bounded by the concentric circles of radii R_0 and r_0 respectively.

$$r < r_0 < R_0 < R.$$

So far the ring-meromorphic case escaped the careful treatment which has been given to the classical meromorphic as well as to the algebroid cases. The customary procedure of letting Φ be set up by a point charge destroys the rotational symmetry in the present case, thus complicating the situation considerably.

The region $G - G_0$ consists of two parts separated by the conductor G_0; one is bounded by the circles of radii R and R_0, let us call it ${}^eG$; the other one, ${}^iG$, by the circles of radii r and r_0. Their capacities are

$${}^eC = \left(\log\frac{R}{R_0}\right)^{-1}, \qquad {}^iC = \left(\log\frac{r_0}{r}\right)^{-1}.$$

Using these designations it is found that the capacity of the condenser $\mathfrak{F}$ is

$$C = {}^eC + {}^iC. \qquad \text{(Capacities connected in parallel.)}$$

Therefore we find for $\Phi(z)$ the following expressions

$$\Phi(z) = \{{}^eC/C\}\log\frac{R}{|z|} \quad \text{for } z \text{ in } {}^eG,$$

$$\Phi(z) = \{{}^iC/C\}\log\frac{|z|}{r} \quad \text{for } z \text{ in } {}^iG,$$

$$\Phi(z) = \{1/C\} \quad \text{for } z \text{ in } G_0,$$

$$\Phi(z) = 0 \quad \text{everywhere else.}$$

The charges carried by elements of arc-length on the various boundaries are

$$d\sigma = \frac{1}{2\pi}\{{}^eC/C\}\,d\varphi \quad \text{on the boundary of } {}^eG,$$

$$d\sigma = \frac{1}{2\pi}\{{}^iC/C\}\,d\varphi \quad \text{on the boundary of } {}^iG.$$

Writing down the relations (2.2) and (4.2) with these specializations it will be seen that all quantities involved show a characteristic pattern: They are the weighted averages of two quantities each having a similar significance, regarding one of the regions ${}^eG$ or ${}^iG$ alone, as the total terms have with regard to the whole region G. The weights are the capacities of the corresponding regions repectively. Thus

$$f(G) = [{}^ef({}^eG);\, {}^if({}^iG)] = \frac{{}^eC\cdot{}^ef + {}^iC\cdot{}^if}{C}. \tag{5.1}$$

For the compensating terms for example we have

$${}^em({}^eG;\alpha) = \frac{1}{2\pi}\int_0^{2\pi} \log \|\alpha x(Re^{i\varphi})\|^{-1}\,d\varphi = m(R;\alpha),$$

$${}^im({}^iG;\alpha) = \frac{1}{2\pi}\int_0^{2\pi} \log \|\alpha x(re^{i\varphi})\|^{-1}\,d\varphi = m(r;\alpha),$$

and similarly for m^0 and Ω^0. For the terms counting multiplicities let N serve as an example:

$${}^eN({}^eG;\alpha) = \sum_{R>|z|>R_0} h(z;\alpha)\log\frac{R}{|z|} + c\log\frac{R}{R_0} = {}^eN(R;\alpha)$$

$${}^iN({}^iG;\alpha) = \sum_{r<|z|<r_0} h(z;\alpha)\log\frac{|z|}{r} + c'\log\frac{r_0}{r} = {}^iN(r;\alpha)$$

where the constants c and c' must be chosen subject only to the condition

$$c + c' = \sum_{r_0\leq|z|\leq R_0} h(z;\alpha).$$

In other words: It is immaterial how the contributions to N, coming from intersections of $\mathfrak{C}$ with α in points over the conductor G_0, are divided up among ${}^eN$ and ${}^iN$. The resulting arbitrariness in the functions

$$(5.2)\qquad \begin{aligned} {}^eT_l(r) &= {}^eN_l(r;\alpha) + m_l(r;\alpha) - m_l^0(r;\alpha) \qquad (r > R_0)\\ {}^iT_l(r) &= {}^iN_l(r;\alpha) + m_l(r;\alpha) - m_l^0(r;\alpha) \qquad (r < r_0)\end{aligned}$$

coincides with the one which the very structure of any expression (5.1) allows its constituents ${}^ef$ and ${}^if$. These components are not uniquely determined by f, for any pair

$${}^ef + c/{}^eC; \qquad {}^if - c/{}^iC \qquad (c = \text{const.})$$

will perform equally well in their stead. The condition $f(G) > 0$ is satisfied by all quantities entering into the formulae of the first and second main theorems. We remark that under these circumstances a constant c can be found such that also

$$^{e}f^{*} = {}^{e}f + c/{}^{e}C > 0, \qquad {}^{i}f^{*} = {}^{i}f - c/{}^{i}C > 0.$$

Disregarding the positive denominator we are given two functions $f(x)$ and $g(y)$ such that for all values of x and y for which they are explained we have $f(x) + g(y) > 0$. We then have to find a constant c such that also

$$f^{*}(x) = f(x) + c > 0; \qquad g^{*}(y) = g(y) - c > 0.$$

Denoting by f and g the greatest lower bounds of $f(x)$ and $g(y)$ respectively, we see that $c = \frac{1}{2}(g - f)$ certainly has the required property, for

$$f^{*}(x) \geqq \tfrac{1}{2}(f + g) \geqq 0, \qquad g^{*}(y) \geqq \tfrac{1}{2}(f + g) \geqq 0,$$

where never more than one of the equality signs can possibly hold in any given case.

The specializations $r = r_0$ and $R = R_0$ respectively show that the first two main theorems for ring-meromorphic curves, being originally relations between weighted averages, permit the following formulation for the quantities in their unaveraged state: The functions (5.2), defined for sufficiently great and sufficiently small values of their arguments respectively, are independent of the particular plane α. They define in pairs the orders

$$T_l(G) = [{}^{e}T_l(R);\ {}^{i}T_l(r)]$$

of a ring-meromorphic curve in the sense that two such pairs are equivalent:

$$[{}^{e}T(R);\ {}^{i}T(r)] \sim [{}^{e}T^{*}(R);\ {}^{i}T^{*}(r)]$$

if

$$\begin{aligned} {}^{e}T^{*}(R) &= {}^{e}T(R) + c \log R + O(1), \\ {}^{i}T^{*}(r) &= {}^{i}T(r) + c \log r + O(1). \end{aligned}$$

With this same convention the second main theorem appears in the form

$$(5.3) \quad \begin{aligned} {}^{e}V_l(r) + \{{}^{e}T_{l+1}(r) - {}^{e}T_l(r) + {}^{e}T_{l-1}(r)\} &= \Omega_l(r) - \Omega_l(R_0) \quad \text{for } r > R_0, \\ {}^{i}V_l(r) + \{{}^{i}T_{l+1}(r) - {}^{i}T_l(r) + {}^{i}T_{l-1}(r)\} &= \Omega_l(r) - \Omega_l(r_0) \quad \text{for } r < r_0. \end{aligned}$$

We finally remark that, since

$$^{e}m^{0}({}^{e}G; \alpha) = m(R_0\,; \alpha) \sim 0; \qquad {}^{i}m^{0} = m(r_0\,; \alpha) \sim 0$$

we have also

$$m^{0}(G; \alpha) = [m(R_0\,; \alpha); m(r_0\,; \alpha)] \sim 0$$

and similarly

$$\Omega_l^{0}(G) = [\Omega_l(R_0); \Omega_l(r_0)] \sim 0$$

allowing us to replace the equality signs in (5.2) and (5.3) respectively by equivalences and dropping the terms m_l^0 and Ω_l^0.

II. The Third Main Theorem

6. General Relations

Some of the formulae, leading up to the estimates which finally culminate in the defect relations, can be derived in the general case. To do this is the aim of this first section.

Again we consider the function

$$w(\mathfrak{p}) = (\alpha x(\mathfrak{p}))/(\beta x(\mathfrak{p})) = w = w_1/w_2$$

which is meromorphic on $\mathfrak{F}$. We plot its values w on the w-sphere of diameter 1 into which the w-plane passes by stereographic projection. Let us denote by

$$d\tau_w = \frac{dw\,d\bar{w}}{(1 + w\bar{w})^2}$$

its surface element. $w(\mathfrak{p})$ maps the surface $\mathfrak{F}$ in a one-to-one fashion upon a Riemann surface $\mathfrak{F}_w$ covering the w-sphere. Consider now the integral

$$\frac{1}{2\pi}\int_{\mathfrak{F}} \Phi(\mathfrak{p}) \frac{|w'|^2}{(1 + w\bar{w})^2}\,dt\,\overline{dt} \tag{6.1}$$

extended over the whole surface $\mathfrak{F}$, where t is the local uniformizing-parameter and the prime indicates differentiation with respect to t. The differential $|w'|^2\,dt\,\overline{dt}$ is independent of the particular choice of t. Assume that on the w-sphere there is a uniform charge of density 1 and that each point on $\mathfrak{F}_w$ carries the charge of its trace on the w-sphere. Then, if $\mathfrak{q}$ is the image on $\mathfrak{F}_w$ of $\mathfrak{p}$ on $\mathfrak{F}$,

$$\Phi(\mathfrak{p}) \frac{|w'|^2}{(1 + w\bar{w})^2}\,dt\,\overline{dt}$$

is the work required to transport the charge of the surface element of $\mathfrak{F}_w$ at $\mathfrak{q}$ along any path on $\mathfrak{F}$ from infinity into the point $\mathfrak{p}$ against the potential Φ. The integral (6.1) is therefore $1/2\pi$ times the work required to assemble the charge of $\mathfrak{F}_w$ on $\mathfrak{F}$ in such a way that each point carries the charge of its image. On the other hand

$$\sum \Phi(\mathfrak{p})h(\mathfrak{p};\, \alpha_i w_2 - \beta_i w_1)\,d\tau_w$$

the sum extending over all points $\mathfrak{p}$ of $\mathfrak{F}$ is the work required to assemble in the proper places on $\mathfrak{F}$ the charges carried by the elements of $\mathfrak{F}_w$ above the surface element $d\tau_w$ of the w-sphere at the point $w = w_1/w_2$. Hence

$$\frac{1}{2\pi}\int_{\mathfrak{F}} \Phi(\mathfrak{p}) \frac{|w'|^2}{(1 + w\bar{w})^2}\,dt\,\overline{dt} = \frac{1}{2\pi}\int \sum \Phi(\mathfrak{p})h(\mathfrak{p};\, \alpha_i w_2 - \beta_i w_1)\,d\tau_w$$

where the second integral is to be extended over the whole w-sphere. Following the argument previously employed[7] we take the average $\mathfrak{M}_{\alpha\beta}$[8] on both sides and replace w and w' by their expressions in terms of the $x_i(\mathfrak{p})$. We then obtain

$$(6.2)\qquad \frac{1}{2\pi}\int_{\mathfrak{F}} \Phi(\mathfrak{p})\mathfrak{M}_{\alpha\beta}(Q^2)\, dt\, d\bar{t} = \frac{1}{2\pi}\int \mathfrak{M}_{\alpha\beta}N(G;\alpha_i w_2 - \beta_i w_1)\, d\tau_w$$

with

$$Q = \frac{\left|\sum\limits_{i<j} [\alpha\beta]_{ij}[xx']_{ij}\right|}{\left|\sum_0^k \alpha_i x_i\right|^2 + \left|\sum_0^k \beta_i x_i\right|^2}.$$

Next we normalize the homogeneous form w_1/w_2 of w so that

$$\bar{w}_1 w_1 + \bar{w}_2 w_2 = 1.$$

Consequently the vectors

$$\xi_i = \alpha_i w_2 - \beta_i w_1, \qquad \eta_i = \alpha_i \bar{w}_1 + \beta_i \bar{w}_2,$$

form a unitary pair if α and β do, and in particular we have

$$\sum_0^k |\xi_i|^2 = \bar{w}_1 w_1 + \bar{w}_2 w_2 = 1.$$

Hence replacing α and β in the second integral by ξ and η it becomes by Lemma 1. M.C. pg. 519

$$\frac{1}{2\pi}\int \mathfrak{M}_\xi N(G;\xi)\, d\tau_w = \tfrac{1}{2}T(G)$$

and therefore

$$\frac{1}{2\pi}\int_{\mathfrak{F}} \Phi(\mathfrak{p})\mathfrak{M}_{\alpha\beta}(Q^2)\, dt\, d\bar{t} = \tfrac{1}{2}T(G).$$

If we introduce upon the w-sphere a charge of density $\mu_w > 0$ in each point, varying with w, then the charge carried by an element of surface $d\tau_w$ will be $\mu d\tau_w$. The density μ may also depend on α and β, but if it does we presuppose that it be a homogeneous function of the combinations ξ_i. Under these conditions (6.2) is replaced by

$$(6.3)\qquad \frac{1}{2\pi}\int_{\mathfrak{F}} \Phi(\mathfrak{p})\mathfrak{M}_{\alpha\beta}\{Q^2\mu(\Xi_i)\}\, dt\, d\bar{t} = \tfrac{1}{2}\mathfrak{M}_\xi\{N(G;\xi)\mu(\xi)\}$$

where the Ξ_i arise from $\xi_i = \alpha_i w_2 - \beta_i w_1$ if we replace w_1 and w_2 by their values $\sum_0^k \alpha_j x_j(\mathfrak{p})$ and $\sum_0^k \beta_j x_j(\mathfrak{p})$ respectively:

$$\Xi_i = \sum_j (\alpha_i\beta_j - \alpha_j\beta_i)x_j(\mathfrak{p}).$$

[7] M C. pp. 528–530.

[8] M C. p. 518.

The value of either integral equals again the average amount of work required to assemble the total charge of an $\mathfrak{F}_w$ on $\mathfrak{F}$ in such a way that each point carries the charge of its image. With the aid of

$$N(G;\, \xi) \leqq T(G) + m^0(G;\, \xi)$$

we obtain from (6.3)

$$\frac{1}{2\pi}\int_{\mathfrak{F}} \Phi(\mathfrak{p})\mathfrak{M}_{\alpha\beta}\{Q^2\mu(\Xi_i)\}\, dt\, d\bar{t} \leqq cT(G) + c'(G) \tag{6.4}$$

where

$$c = \mathfrak{M}_\xi \mu(\xi), \qquad c' = \mathfrak{M}_\xi\{m^0(G;\, \xi)\mu(\xi)\}.$$

Again the difficulty arises that in general c' will depend on G. But c' is

$$c' = \int_{\Gamma_0} \mathfrak{M}_\xi\{\mu(\xi) \log \|\, \xi x\,\|^{-1}\}\, d\sigma$$

and therefore bounded if the mean value appearing under the integral sign is a continuous function of $x_0, x_1, \cdots, x_k$. (In the cases of meromorphic, algebroid, and ring-meromorphic curves this difficulty does not exist: c' is finite with c.)

Let us choose the density of charge on the w-sphere in a fashion which closely resembles the customary choices made in the theory of meromorphic functions.

$$\mu(\xi) = \prod_{\nu=1}^{q} \|\, a^{(\nu)}\xi\,\|^{-2\lambda_\nu} \tag{6.5}$$

where the product extends over q arbitrary points $a^{(1)}, a^{(2)}, \cdots, a^{(q)}$ in $\mathfrak{R}$. The exponents, which are assumed to be real non-negative numbers, must be determined such that $M_\xi\mu(\xi)$ is finite. With such a determination of the λ_ν the particular form (6.5) of $\mu(\xi)$ will also assure the finiteness of c'. Independently of the curve $\mathfrak{C}$ we have the relation

$$\mathfrak{M}_{\alpha\beta}\{\log Q^2 + \log \mu(\Xi_i)\} = \text{const.} + 2\log\{X_2/X_1\} + 2\textstyle\sum_a \lambda_a \log [\![ax]\!]^{-1} \tag{6.6}$$

for the choice (6.5) of $\mu(\xi)$.[9]

7. The Conditions on the λ_ν

Let us denote by $d\omega_\xi$ the volume element of the unitary k-sphere $\mathfrak{S}_k : \rho^2 = \sum_0^k |\, \xi_i\,|^2 = 1$. Then our task is to determine the exponents or—as we shall call them—weights λ_ν in such a manner that the integral

$$J = \int_{\mathfrak{S}_k} \mu(\xi)\, d\omega_\xi$$

[9] For a proof see: M C. pp. 530, 535.

converges. This convergence does not depend upon the nature of the curve $\mathfrak{C}$ but merely on the linear dependence scheme of the arbitrary points $a^{(\nu)}$ in $\mathfrak{R}$. It can be shown by the methods employed before (M. C. pg. 536) that J is finite for $\lambda_\nu < 1$, $(\nu = 1, 2, \cdots, q)$ if the points $a^{(\nu)}$ satisfy no accidental linear relations, i.e. if any $(k + 1)$ of them are linearly independent. From this restriction we wish to free ourselves.

The integral whose convergence we are to investigate can be written in the form

$$J = \int d\omega_\xi/\Pi(\xi), \qquad \Pi(\xi) = \prod_{\nu=1}^{q} |(a^{(\nu)}\xi)|^{2\lambda_\nu}$$

where the product $\Pi(\xi)$ is formed over a finite number of linear forms $(a\xi)$ whose coefficients a_i are normalized so as to satisfy the condition $\sum_0^k |a_i|^2 = 1$. Each of them is furthermore provided with a weight $\lambda \geqq 0$.

We interprete the ξ_i as components of a vector in a complex $(k + 1)$-dimensional vector space. The sum $\sum \lambda$ extended over all planes $(a\xi)$ which contain a given l-dimensional linear submanifold S_l of that vector space shall be called the load carried by S_l.

First we investigate the somewhat simpler space integral

$$\int dV_\xi/\Pi(\xi) \qquad \text{extended over } \rho^2 \leqq r_0^2. \tag{7.1}$$

Before proceeding any further it is necessary to express the volume element dV_ξ of the real $(2k + 2)$-dimensional space with complex coordinates ξ_i. What we are seeking is the generalization of the surface element $dz\,d\bar{z}$ employed instead of $dx\,dy$ when the (x, y)-plane is described by the complex coordinate $z = x + iy$.

Let our space be spanned by the $2k + 2$ real basis vectors

$$\mathfrak{e}_0, \mathfrak{e}_0', \mathfrak{e}_1, \mathfrak{e}_1', \cdots, \mathfrak{e}_k, \mathfrak{e}_k'.$$

We split our coordinates ξ_i into real and imaginary parts

$$\xi = x + iy \qquad (i = \sqrt{-1})$$

as suggested by the two-dimensional case. An arbitrary line element

$$\mathfrak{d} = (d\xi_0, d\xi_1, \cdots, d\xi_k)$$

then has the components

$$dx_0, dy_0, dx_1, dy_1, \cdots, dx_k, dy_k$$

with respect to the above vector basis. Let the line element

$$(id\xi_0, id\xi_1, \cdots, id\xi_k)$$

with the corresponding complex components be called $\mathfrak{d}'$. We note that the transition from $\mathfrak{d}$ to $\mathfrak{d}'$ is invariant under an analytic coordinate transformation. The real components of $\mathfrak{d}'$ are

$$-dy_0, dx_0, -dy_1, dx_1, \cdots, -dy_k, dx_k.$$

If we have $(k + 1)$ line elements

$$\begin{aligned} \mathfrak{d}_0 &= (d\xi_0, d\xi_1, \cdots, d\xi_k) \\ \mathfrak{d}_1 &= (\delta\xi_0, \delta\xi_1, \cdots, \delta\xi_k) \\ &\vdots \end{aligned}$$

linearly independent in the complex sense then $\mathfrak{d}_0, \mathfrak{d}_0', \cdots, \mathfrak{d}_k, \mathfrak{d}_k'$, are linearly independent in the real sense. As volume element we use the parallelopiped spanned by them:

$$dV_\xi = \begin{vmatrix} dx_0 & dy_0 & dx_1 & dy_1 & \cdots \\ -dy_0 & dx_0 & -dy_1 & dx_1 & \cdots \\ \delta x_0 & \delta y_0 & \delta x_1 & & \\ \vdots & \vdots & \vdots & & \end{vmatrix}.$$

From their definitions it follows that

$$\begin{aligned} \mathfrak{d} + i\mathfrak{d}' &= \overline{d\xi}_0(\mathfrak{e}_0 + i\mathfrak{e}_0') + \cdots + \overline{d\xi}_k(\mathfrak{e}_k + i\mathfrak{e}_k'), \\ \mathfrak{d} - i\mathfrak{d}' &= d\xi_0(\mathfrak{e}_0 - i\mathfrak{e}_0') + \cdots + d\xi_k(\mathfrak{e}_k - i\mathfrak{e}_k'), \end{aligned}$$

hence the desired volume element appears in the form

$$dV_\xi = \bar{\Delta}\Delta$$

where

$$\Delta = \begin{vmatrix} d\xi_0 & d\xi_1 & \cdots & d\xi_k \\ \delta\xi_0 & \delta\xi_1 & \cdots & \delta\xi_k \\ \vdots & \vdots & & \vdots \end{vmatrix}$$

LEMMA 1. The integral (7.1) is convergent if each linear vector manifold of dimension $(k + 1 - l)$ carries a load

$$\Lambda_l < l. \qquad (l = 1, 2, \cdots, k + 1).$$

The proof shall be carried through by means of an induction with respect to the number of dimensions.

Let us assume that among the planes $(a^{(\nu)}\xi) = 0$ there are at least $(k + 1)$ linearly independent ones. Should this not be the case we can introduce additional linear forms with the weights $\lambda = 0$.

The set $\mathfrak{A}$ of planes a: $(a\xi) = 0$, $\sum_0^k |a_i|^2 = 1$, in the $(k + 1)$-dimensional vector space $\mathfrak{E}_{k+1}$ shall shortly be referred to as the configuration $\{a\}$. The intersection of any k linearly independent planes $(a^{(\nu)}\xi) = 0$ shall be called a vertex $\mathfrak{z}$. It is our intention to find corresponding to every configuration of this kind a cell-division of the ξ-space such that each vertex $\mathfrak{z}$ is contained in a cell $\mathfrak{Z}$ which is cut by no other planes of $\mathfrak{A}$ except those which pass through $\mathfrak{z}$. This can be accomplished inductively, for any plane a of $\mathfrak{A}$ is an $\mathfrak{E}_k$ and its in-

tersections $\bar{\mathfrak{b}}$ with all planes of $\mathfrak{A}$ different from a form a configuration in this lower-dimensional space. Assume therefore that we have found a cell-division with the desired property for the configuration $\{\bar{\mathfrak{b}}\}$ in this $\mathfrak{E}_k$. Let $\mathfrak{Z}_k$ be one of its cells with vertex $\mathfrak{z}$, a flat cell, so to speak, which we have to extend into space. To this end we perform a unitary transformation

$$(\xi_0, \xi_1, \cdots, \xi_k) \to (\zeta_0, \zeta_1, \cdots, \zeta_k)$$

such that $(a\xi) = \zeta_0$, i.e. such that the ζ_0-axis is orthogonal to the plane $(a\xi) = 0$. Then we define

$$(\zeta_0, \zeta_1, \cdots, \zeta_k) \text{ is in } \mathfrak{Z}_{k+1} \text{ if}$$

$$\text{D 1.)} \quad (0, \zeta_1, \zeta_2, \cdots, \zeta_k) \text{ is in } \mathfrak{Z}_k .$$

$$\text{D 2.)} \quad |\zeta_0| \leqq |(b\zeta)| \text{ for all } b \text{ in } \mathfrak{A}.$$

This inductive definition begins with

$$\mathfrak{z} = \mathfrak{Z}_1 .$$

Let us call $\mathfrak{z}$ the *center* and $\zeta_0 = 0$ the *base* of $\mathfrak{Z}_{k+1}$. The cells thus defined have the following properties:

(1) They are cones in the sense that $(\zeta_0, \zeta_1, \cdots, \zeta_k)$ in $\mathfrak{Z}_{k+1}$ implies $(\lambda\zeta_0, \cdots, \lambda\zeta_k)$ in $\mathfrak{Z}_{k+1}$ for any complex constant λ.

(2) They are closed, as evidenced by the equality sign in D 2.).

(3) They cover the whole space: Let $\zeta^{(0)}$ be any point of our space then there will be an $|(a\zeta^{(0)})|$ such that none of the expressions $|(b\zeta^{(0)})|$ (b in $\mathfrak{A}$, $b \neq a$) have a smaller value, consequently $\zeta^{(0)}$ belongs to a cell with the base a.

(4) A cell contains but one vertex $\mathfrak{z}$, namely its center.

If $\mathfrak{z}'$ were in $\mathfrak{Z}$ then the k linearly independent planes defining $\mathfrak{z}'$ would also have to go through $\mathfrak{z}$ which is impossible unless $\mathfrak{z}' \equiv \mathfrak{z}$. Finally we have

(5) A plane b of $\mathfrak{A}$ has no point other than the origin $(0, 0, \cdots, 0)$ in common with a cell unless it passes through its center.

Assume this to be true for the flat cells $\mathfrak{Z}_k$ in $\mathfrak{E}_k$ and the manifolds $\bar{\mathfrak{b}}$ of the configuration $\{\bar{\mathfrak{b}}\}$. Let b be a plane such that for some point $\zeta^{(0)} \neq (0, \cdots, 0)$ in $\mathfrak{Z}_{k+1}$ we have $(b\zeta^{(0)}) = 0$. Either b is the plane $\zeta_0 = 0$ or it is not. In either case b is seen to pass through $\mathfrak{z}$. In the first one this is evident; in the second one b intersects the plane $\zeta_0 = 0$ in an $\mathfrak{E}_{k-1}$ shortly denoted by $\bar{\mathfrak{b}}$. We remark that it follows from $(b\zeta^{(0)}) = 0$ that also $\zeta_0^{(0)} = 0$ on account of D 2.). Hence the point $\zeta^{(0)}$ is firstly in the flat cell $\mathfrak{Z}_k$ and secondly contained in $\bar{\mathfrak{b}}$; therefore it follows from the induction hypothesis that $\bar{\mathfrak{b}}$ must contain $\mathfrak{z}$, implying the same for b. The contention is evidently true for $k = 1$ since every $\mathfrak{E}_1$ of our configuration is some cell's center, any two of them co-inciding completely as soon as they have a point other than $(0, 0, \cdots, 0)$ in common.

For any plane b of $\mathfrak{A}$ not passing through $\mathfrak{z}$ we even have the quantitative estimate

$$(7.2) \qquad |(b\zeta)| \geqq \delta(\textstyle\sum_0^k |\zeta_i|^2)^{\frac{1}{2}}$$

with a positive constant δ for all points ζ of a cell $\mathfrak{Z}$ with center $\mathfrak{z}$. This follows at once from the fact that the function $(\sum_0^k |\zeta_i|^2)^{\frac{1}{2}} : |(b\zeta)|$ is a bounded function of its arguments on the closed set $\mathfrak{Z}$. The quantity

$$\beta = |(b\mathfrak{z}_i^{(0)})| : (\textstyle\sum_0^k |\mathfrak{z}_i^{(0)}|^2)^{\frac{1}{2}}$$

has the same value for all points $\mathfrak{z}^{(0)}$ of $\mathfrak{z}$ (The "distance" of the vertex $\mathfrak{z}$ from the plane b): If b does not pass through $\mathfrak{z}$ we have $\beta > 0$ and can derive the explicit estimate

$$\delta \geqq 2^{-k}\beta. \qquad \text{(See Appendix)}$$

If we unite all cells $\mathfrak{Z}_{k+1}$ with center $\mathfrak{z}$ we obtain the star $\mathfrak{Z}_{k+1}^*$. All points of $\mathfrak{z}$ with exception of $(0, 0, \cdots, 0)$ will be interior points of $\mathfrak{Z}_{k+1}^*$. The estimate (7.2) remains true for all points of $\mathfrak{Z}_{k+1}^*$ if I choose for δ the smallest one of all those δ's which correspond to the cells combined into $\mathfrak{Z}_{k+1}^*$.

We designate by $\mathfrak{Z}^{(s)}$ the intersection of $\sum_0^k |\zeta_i|^2 \leqq r_0^2$ and $\mathfrak{Z}_{k+1}^* : \mathfrak{Z}^{(s)} = \{\sum_0^k |\zeta_i|^2 \leqq r_0^2\} \frown \mathfrak{Z}_{k+1}^*$. The region over which (7.1) is to be integrated consists of a finite number of such stumps $\mathfrak{Z}^{(s)}$. In order to evaluate the part of (7.1) extending over any one of them we shall specify the coordinate system $(\zeta_0, \zeta_1, \cdots, \zeta_k)$ more accurately than has been done so far. We choose a cell $\mathfrak{Z}_{k+1}$ of the star $\mathfrak{Z}_{k+1}^*$: Let its base be $\mathfrak{E}_k$. The corresponding flat cell will again have a base $\mathfrak{E}_{k-1}$, and so forth. Now we determine the $\zeta_0, \zeta_1, \cdots, \zeta_k$ such that

$$\begin{aligned} \mathfrak{E}_k \quad & \text{is } \zeta_0 = 0, \\ \mathfrak{E}_{k-1} \ & \text{is } \zeta_0 = 0,\ \zeta_1 = 0, \\ & \ \vdots \\ \mathfrak{z} = \mathfrak{E}_1 \quad & \text{is } \zeta_0 = 0,\ \zeta_1 = 0, \cdots, \zeta_{k-1} = 0. \end{aligned}$$

We demand furthermore that the ζ_i arise from the ξ_i by a unitary transformation. This determines them uniquely to within factors of absolute value 1. We refer to the system of the ζ_i-axes as the *frame work* spanning the cell $\mathfrak{Z}_{k+1}$. The coordinate system which we introduce in the star $\mathfrak{Z}_{k+1}^*$ is the frame work of any one of its cells.

Let $(b\zeta) = 0$ be some plane b of $\mathfrak{A}$ not going through $\mathfrak{z}$. Then the planes $\zeta_0 = 0, \zeta_1 = 0, \cdots, \zeta_{k-1} = 0, (b\zeta) = 0$, form a system of $(k + 1)$ linearly independent planes in terms of which any plane of our configuration $\{a\}$ can be expressed in linear fashion. We denote by $c_0, c_1, \cdots, c_{k-1}$, the coefficients of $\zeta_0, \zeta_1, \cdots, \zeta_{k-1}$, and by c_k the coefficient of $(b\zeta)$ in this representation:

$$(a\zeta) \equiv \textstyle\sum_0^{k-1} c_i\zeta_i + c_k(b\zeta).$$

Furthermore we have from (7.2) that $|b_k| \geqq \delta$.

Now we perform the substitution

$$\begin{aligned} (b\zeta) &= z \\ \zeta_0 &= \eta_0 z \\ \zeta_1 &= \eta_1 z \\ &\vdots \\ \zeta_{k-1} &= \eta_{k-1} z\,. \end{aligned}$$

The successive transformations

$$(\xi_0, \xi_1, \cdots, \xi_k) \to (\zeta_0, \zeta_1, \cdots, \zeta_{k-1}, (b\zeta)) \to (z, \eta_0, \cdots, \eta_{k-1})$$

bring our volume element into the form

$$dV_\xi = |b_k|^{-2}\, dV_\zeta = |b_k|^{-2}(z\bar{z})^k\, dz\, d\bar{z}\, dV_\eta$$

and our integral can now be written as

$$\int dV_\xi/\Pi(\xi) = |b_k|^{-2} \int (z\bar{z})^{k-\Lambda}\, dz\, d\bar{z} \cdot \prod_1^q |c_0^{(\nu)}\eta_0 + \cdots + c_{k-1}^{(\nu)}\eta_{k-1} + c_k^{(\nu)}|^{-2\lambda_\nu}\, dV_\eta$$

where $\Lambda = \sum_1^q \lambda_\nu$ is the total load carried by the origin. Both integrations are to be extended over the stump $\mathfrak{Z}^{(s)}$.

The linear forms in the denominator of the second factor will be divided into two classes—the first one containing the forms corresponding to planes that do—the second one those corresponding to planes that do not pass through $\mathfrak{z}$. For any one of the latter kind we have

$$\left| \sum_0^{k-1} c_i\zeta_i + c_k(b\zeta) \right| \geqq \delta' \left\{ \sum_0^k |\zeta_i|^2 \right\}^{1/2} \geqq \delta' |z| \qquad (c_k \neq 0)$$

hence

$$|c_0\eta_0 + \cdots + c_{k-1}\eta_{k-1} + c_k| \geqq \delta' \qquad (\delta' = \text{const.} > 0).$$

This permits us to write

$$(7.3) \qquad \int dV_\xi/\Pi(\xi) \leqq c \int (z\bar{z})^{k-\Lambda}\, dz\, d\bar{z} \cdot \prod_\mu |c_0^{(\mu)}\eta_0 + \cdots + c_{k-1}^{(\mu)}\eta_{k-1}|^{-2\lambda_\mu}\, dV_\eta$$

where $c > 0$ is some constant, and the product in the denominator of the second term is to be extended over all those planes $a^{(\mu)}$ of $\mathfrak{A}$ which pass through $\mathfrak{z}$. Again both integrals are to be extended over $\mathfrak{Z}^{(s)}$.

Concerning the limits of integration in the new variables we note that from $\sum_0^k |b_i|^2 = 1$ and $\sum_0^k |\zeta_i|^2 \leqq r_0^2$ follows $|z| \leqq r_0$ since

$$\frac{|(b\zeta)|^2}{\sum_0^k |\zeta_i|^2} = \frac{|z|^2}{\sum_0^k |\zeta_i|^2} \leqq 1.$$

On the other hand we have from (7.2)

$$|z| \geqq \delta\left\{\sum_0^k |\zeta_i|^2\right\}^{1/2} \geqq \delta\left\{\sum_0^{k-1} |\zeta_i|^2\right\}^{1/2},$$

finally yielding

$$|\eta_0|^2 + |\eta_1|^2 + \cdots + |\eta_{k-1}|^2 \leqq \delta^{-2}.$$

In other words: The stump $\mathfrak{Z}^{(\sigma)}$ is contained in the region described by

$$|z| \leqq r_0, \qquad \sum_0^{k-1} |\eta_i|^2 \leqq \delta^{-2}.$$

Integrating (7.3) over this region rather than $\mathfrak{Z}^{(\sigma)}$ will increase the integral. We introduce polar coordinates for the complex variable

$$z = \rho e^{i\varphi} \qquad (i = \sqrt{-1})$$

and thus obtain

$$\int_{\mathfrak{Z}^{(\sigma)}} \frac{dV_\xi}{\Pi(\xi)} \leqq c \int_0^{r_0} \int_0^{2\pi} \rho^{2k+1-2\Lambda}\, d\rho\, d\varphi \cdot \int \frac{dV_\eta}{\Pi(\eta)}$$

where the last integral is to be extended over the sphere

$$\sum_0^{k-1} |\eta_i|^2 \leqq \delta^{-2}.$$

The convergence of the first factor is guaranteed by the condition $\Lambda < k + 1$. The convergence of the second factor is the contention of Lemma 1 with k replaced by $(k - 1)$, i.e. the induction hypothesis.

We anchor this induction at $k = 1$ for which the lemma is evidently true.

The integral J converges under the reduced assumption which arises from the one of Lemma 1 by dropping the restriction that the origin carries a load $< k + 1$. This is shown by modifying the first step of our proof, integrating (7.1) over a spherical shell

$$0 < r_1 \leqq \rho \leqq r_0 \tag{7.4}$$

rather than over a solid sphere.

$$\int dV_\xi/\Pi(\xi) = J \int_{r_1}^{r_0} \rho^{2k+1-2\Lambda}\, d\rho,$$

where the first integral is extended over the shell (7.4). Concerning the limits of integration we note that the intersection of $\mathfrak{Z}_{k+1}^*$ with the shell is contained in the region

$$\delta r_1 \leqq |z| \leqq r_0, \qquad \sum_0^{k-1} |\eta_i|^2 \leqq \delta^{-2}$$

Therefore the part of J extending over the intersection of $\mathfrak{Z}_{k+1}^*$ with the unitary k-sphere $\mathfrak{S}_k$ will be less than or equal to

$$\text{const} \int dV_\eta/\Pi(\eta) \text{ extended over } \sum_0^{k-1} |\eta_i|^2 \leqq \delta^{-2}.$$

Thus the proof of what we maintained is reduced to an application of Lemma 1 for $(k - 1)$ instead of k. Returning to the projective way of expression and considering the ξ_i as homogeneous plane-coordinates, the $a_i^{(\nu)}$ as homogeneous point-coordinates, we obtain the

LEMMA 2. The integral J converges if each $(l-1)$-dimensional linear subspace of the projective k-space $(1 \leqq l \leqq k)$ carries a load $\Lambda_l < l$. (The load Λ_l of a given subspace is the sum of the weights of the points $a^{(\nu)}$ contained therein.)

The proof of this lemma indicates the reason why no restriction need be made upon the total load carried by the whole space. It is the load of the origin in the ξ-space and of no influence upon the convergence of the integral J since the latter extends only over points $\sum_0^k |\xi_i|^2 = 1$.

For the choice (6.5) of $\mu(\xi)$ the average $\mathfrak{M}_\xi \mu(\xi)$ will remain finite if the weights λ_ν are so chosen that any given $(l-1)$-spread in $\mathfrak{R}$ $(l = 1, \cdots, k)$ carries a load $\Lambda_l < l$. This, as we remarked once before, will imply the finiteness of

$$c' = \mathfrak{M}_\xi \{ m^0(G; \xi) \mu(\xi) \}$$

as well. Let us assume therefore that in the following special cases the λ_ν always comply with the conditions of Lemma 2.

8. The Third Main Theorem for Algebroid Curves

Once more we turn to the case where $\mathfrak{F}$ is an n-sheeted unbounded covering surface of the finite z-plane. We take the set of $(k+1)$ functions defining a realization of $\mathfrak{F}$ and substitute them for the x_i in (6.6). The prime in

$$X_2^2 = \sum_{i<j} | [xx']_{ij} |^2$$

we interpret as differentiation with respect to z. Multiplying through by $d\sigma = (1/2\pi)\, d\varphi$ and integrating over the boundary of G we shall obtain, if we make use of the concavity of the logarithmic function,

$$2\Omega_1(r) + 2\textstyle\sum_\alpha \lambda_\alpha m^*(r; a) + \text{const} \leqq \log \left\{ \dfrac{1}{2\pi n} \displaystyle\int_0^{2\pi n} \mathfrak{M}_{\alpha\beta} \{ Q^2 \mu(\Xi_i) \} \, d\varphi \right\} = \Theta(r).$$

We indicate the rotational symmetry of the potential Φ by writing

$$\Phi(z) = \Phi(\rho) \quad \text{when} \quad |z| = \rho.$$

Then (6.3) yields for $\Theta(\rho)$ the relation

$$\int_0^r \Phi(\rho) e^{\Theta(\rho)} \rho \, d\rho = \int_{r_0}^r \frac{dr}{r} \int_0^r e^{\Theta(\rho)} \rho \, d\rho \leqq cT(r) + c',$$

denoted shortly by

$$\Theta(r) = \omega(T(r)).$$

From this it follows in customary fashion that

$$\Theta(r) < \kappa \log T(r) - 2 \log r,$$

an inequality holding "almost everywhere," i.e. outside of certain intervals I_r

whose logarithmic measure is finite: $\int_{I_r} \frac{dr}{r} < \infty$.[10] κ is an arbitrary constant $\kappa > 1$, a meaning which this symbol shall retain throughout this chapter.

Formulating the resulting relations at once for higher l we have

$$\Omega_l(r) + \textstyle\sum_a \lambda_a m_l^*(r; a) = \frac{1}{2}\omega(T_l(r))$$

or, in the form of an inequality holding almost everywhere,

$$\Omega_l(r) + \textstyle\sum_a \lambda_a m_l^*(r; a) < \frac{\kappa}{2} \log T_l(r) - \log r$$

with weights λ_a attached to the points a and so chosen that the total load carried by any one $(h - 1)$-spread $(1 \leqq h \leqq k_l)$ in $\mathfrak{R}_l$ is less than h.

These so-called defect relations constitute the third main theorem, holding in this form regardless of any accidental linear relations between the points over which the sum on the left is to be extended, i.e. in particular those points for which m^* is not ~ 0. (Exceptional points)

The methods[11] used for the derivation of estimates of the subsequent nature yield again

$$\Omega_l(r) < \frac{\kappa}{2} \log T(r)$$

hence

$$T_l(r) < lT(r) + \kappa \frac{l(l+1)}{4} \log T(r)$$

and, introducing the symbols

$$\Delta_l(r) = V_l(r) + \sum_a \lambda_a m_l^*(r; a),$$

we finally obtain

$$\Delta_l(r) < \frac{k+1}{k+1-l} T(r) + \frac{k(k+1)}{2(k+1-l)} J(r) + \kappa \frac{k(k+1)}{4(k+1-l)} \log T(r) \tag{8.1}$$

holding almost everywhere. To complete our results we have to give an estimate for the density $J(r)$ of branchpoints of $\mathfrak{F}$ over the z-plane. E. Ullrich proves[12] that for the characteristic $T^*(r)$ of any algebroid function $w = f(z)$, one-valued on $\mathfrak{F}$,

$$J(r) \leqq 2(n-1)T^*(r).$$

This characteristic $T^*(r)$, as defined by E. Ullrich, is in the sense of equivalence equal to the order of the algebroid curve $w_1/w_2 = w$ defined in the complex

[10] M C. pp. 527–528.
[11] M C. pp. 532–533.
[12] l. c. pp. 209–210.

one-dimensional projective space $\{w_1, w_2\}$ as a realization of $\mathfrak{F}$. Such a curve will certainly be defined by projecting our curve $\mathfrak{C}$ from some $(k - 2)$-dimensional linear subspace of $\mathfrak{R}$. Application of (3.5) shows us that

$$T^*(r) \leqq T(r) + \text{const.}$$

since in our case

$$\int_{\Gamma_0} \log (X : X')\, d\sigma = \text{const.}$$

The resulting relation

$$J(r) \leqq 2(n - 1)T(r) + \text{const.} \tag{8.2}$$

shows that not only the level of transcendency but also the degree of ramification of an algebroid curve is set by the first order $T(r)$. By means of (8.2) we obtain from (8.1) the relations

$$V_l(r) + \textstyle\sum_a \lambda_a m_l^*(r; a) < \dfrac{(k+1)(nk - k + 1)}{k + 1 - l}\, T(r) + S(r)$$

holding almost everywhere with

$$S(r) = O\,(\log rT(r)).$$

Thus the defect relation appear in the form that was given them by R. Nevanlinna: $k = 1$; $n = 1$, and E. Ullrich: $k = 1$.

9. The Third Main Theorem for Ring-meromorphic Curves

Instead of the function $x_i(\mathfrak{p})$, meromorphic on the doubly punctured plane $\mathfrak{F}$, which define its realization $\mathfrak{C}$ we use the functions $x_i(z)$ defined for z ranging over that part of the z-plane which is covered by $\mathfrak{F}$. For the choice of G_0 and the exhausting sequence $\{G_{R,r}\}$ made heretofore the potential $\Phi(z)$ is a function of $|z| = \rho$ rather than of z:

$$\Phi_G(z) = \Phi(R, r; \rho).$$

From (6.6) it follows that for

$$\Theta(\rho) = \log\left\{\frac{1}{2\pi}\int_0^{2\pi} \mathfrak{M}_{\alpha\beta}\{Q^2\mu(\Xi_i(\rho e^{i\varphi}))\}\, d\varphi\right\}$$

we have

$$2\Omega_1(\rho) + 2\textstyle\sum_a \lambda_a m^*(\rho; a) + \text{const} \leqq \Theta(\rho)$$

or

$$2\Omega_1(G_{R,r}) + 2\textstyle\sum_a \lambda_a m^*(G_{R,r}\,; a) + \text{const} \leqq [\Theta(R); \Theta(r)].$$

On the other hand (6.3) shows that

$$\int_r^R \Phi(R, r; \rho) e^{\Theta(\rho)} \rho\, d\rho \leqq cT(G_{R,r}) + c'$$

with constants c and c'. To set in evidence the fact that the integral on the left also possesses the characteristic pattern of all quantities related to ring-meromorphic curves we write it as

$$\int_r^R \Phi(R, r; \rho)e^{\Theta(\rho)}\rho\, d\rho = [{}^e\Psi(R); {}^i\Psi(r)]$$

with

$${}^e\Psi(R) = \int_{R_0}^R \frac{dR}{R}\int_{R_0}^R e^{\Theta(\rho)}\rho\, d\rho + \log\frac{R}{R_0}\left\{\int_{r_m}^{R_0} e^{\Theta(\rho)}\rho\, d\rho\right\},$$

$${}^i\Psi(r) = \int_r^{r_0} \frac{dr}{r}\int_r^{r_0} e^{\Theta(\rho)}\rho\, d\rho + \log\frac{r_0}{r}\left\{\int_{r_0}^{r_m} e^{\Theta(\rho)}\rho\, d\rho\right\}.$$

Thus we finally obtain

$$[{}^e\Psi(R); {}^i\Psi(r)] \leqq c[{}^eT(R); {}^iT(r)] + c'. \tag{9.1}$$

The arbitrariness

$${}^e\Psi(R) \to {}^e\Psi(R) + c\log\frac{R}{R_0}, \quad {}^i\Psi(r) \to {}^i\Psi(r) - c\log\frac{r_0}{r}$$

finds again an immediate geometric expression in the freedom of choice of the intermediate circle of radius r_m.

Suppose two functions $\Theta(r)$ and $T(r)$ are defined for sufficiently large values of r. Then we agreed[13] to write

$$\Theta(r) = \omega(T(r))$$

if there exists a centrally symmetric potential $U(r)$ due to a distribution of free charge of density $\exp\Theta(r)$:

$$\frac{1}{r}\frac{\partial}{\partial r}\left\{r\frac{\partial U(r)}{\partial r}\right\} = \exp\Theta(r),$$

such that

$$U(r) \leqq cT(r). \qquad (c = \text{const.} > 0)$$

It follows that there exist constants of integration a and b such that

$$U(r) = \int_{r_0}^r \frac{dr}{r}\int_{r_0}^r \exp\Theta(\rho)\rho\, d\rho - a\log r - b$$

satisfies this relation, r_0 being an arbitrary lower bound for the integrations, subject only to the condition that $\Theta(r)$ and $T(r)$ are defined for $r \geqq r_0$.

We note concerning the relation $\Theta = \omega(T)$ that it implies:

(1) $\Theta(r) = \omega(T^*(r))$ for any $T^*(r) = T(r) + c'\log r + c''$ with constants c' and c'', because the additional terms correspond to free charges of density zero.

[13] M C. p. 528.

(2) $\Theta'(r) = \omega(T)$ for any $\Theta' \leqq \Theta$, because

$$U'(r) = \int_{r_0}^{r} \frac{dr}{r} \int_{r_0}^{r} \exp \Theta'(\rho)\rho \, d\rho - a \log r - b \leqq U(r).$$

(3) The existence of constants c' and c'' such that

$$T^*(r) > 0$$

because for

$$T^*(r) = T(r) + \frac{a}{c} \log r + \frac{b}{c}$$

we have

$$cT^*(r) \geqq \int_{r_0}^{r} \frac{dr}{r} \int_{r_0}^{r} \exp \Theta(\rho)\rho \, d\rho > 0.$$

(4) $\Theta(r) < \kappa \log T^*(r)$ almost everywhere if the constants c' and c'' are so chosen as to make $T^*(r) > 0$.

The additional term of the order of $\log r$ has been neglected in the presence of the term $\log T^*$, compared to which it is of importance only in the cases of lowest transcendency.

We extend the meaning of this symbol to functions $\Theta(r)$ and $T(r)$ defined for sufficiently small values of r by writing

$$\Theta(r) = \omega(T(r))$$

if after an inversion on the unit-circle: $r \to 1/r = \tilde{r}$ we have for the functions $\tilde{\Theta}(\tilde{r})$ and $\tilde{T}(\tilde{r})$, defined by

$$\tilde{\Theta}(\tilde{r}) = \Theta(r), \qquad \tilde{T}(\tilde{r}) = T(r),$$

the relation

$$\tilde{\Theta}(\tilde{r}) = \omega(\tilde{T}(\tilde{r})).$$

The application to our case is evident since both ${}^e\Psi$ and ${}^i\Psi$ are centrally symmetric potentials, each due in its region of definition ($r > R_0$, $r < r_0$ respectively) to a distribution of free charges of density $\exp \Theta(r)$. By choosing

$$r = e^{-1} r_0 \text{, while } R \text{ remains variable,}$$

or

$$R = eR_0 \text{, while } r \text{ remains variable}$$

we obtain from (9.1) the inequalities

$${}^e\Psi(R) \leqq c\{{}^eT(R) + \text{const.} \log R + O(1)\}$$

$${}^i\Psi(r) \leqq c\{{}^iT(r) + \text{const.} \log r + O(1)\}$$

with

$$c = \mathfrak{M}_\xi \mu(\xi) > 0.$$

Consequently

$$\Theta(r) = \omega({}^eT(r)), \qquad \Theta(r) = \omega({}^iT(r)),$$

and therefore almost everywhere

$$\Theta(r) < \kappa \log {}^eT^*(r) \qquad \text{for } r > R_0,$$
$$\Theta(r) < \kappa \log {}^iT^*(r) \qquad \text{for } r < r_0,$$

with constants chosen such that ${}^eT^*$ and ${}^iT^*$ are > 0 (see Section 5). Thus the defect relations appear in the form

$$\Omega_l(r) + \textstyle\sum_a \lambda_a m_l^*(r; a) = \frac{1}{2}\omega({}^eT_l(r)) \qquad \text{for } r > R_0,$$
$$\Omega_l(r) + \textstyle\sum_a \lambda_a m_l^*(r; a) = \frac{1}{2}\omega({}^iT_l(r)) \qquad \text{for } r < r_0,$$

or as inequalities holding almost everywhere

$$\Omega_l(r) + \textstyle\sum_a \lambda_a m_l^*(r; a) < \frac{\kappa}{2} \log {}^eT_l^*(r) \qquad \text{for } r > R_0,$$

$$\Omega_l(r) + \textstyle\sum_a \lambda_a m_l^*(r; a) < \frac{\kappa}{2} \log {}^iT_l^*(r) \qquad \text{for } r < r_0.$$

Observing that

$$[\log {}^eT^*; \log {}^iT^*] \leqq \log [{}^eT^*; {}^iT^*]$$

allows us to combine each of the above pairs into a single relation:

$$\Omega_l(G) + \textstyle\sum_a \lambda_a m_l^*(G; a) = \frac{\kappa}{2} \log T_l(G). \qquad (G \equiv G_{R,r})$$

Again we might derive estimates of the kind

$$T_l < lT + S$$

$$V_l < \frac{k+1}{k+1-l} T + S$$

climaxing in

$$V_l + \textstyle\sum_a \lambda_a m_l^* < \frac{k+1}{k+1-l} T + S. \tag{9.2}$$

These relations hold for the averaged quantities as well as for the unaveraged ones, if in the latter case we supplement them by the conventions governing the pair-wise connection between the unaveraged constituents. S stands throughout for a term $O(\log T)$.

Applying (9.2) to a ring-meromorphic function, for instance to one of the quotients x_i/x_0 which define the realization $\mathfrak{C}$, it is seen that any such function must assume every value with the possible exception of at most two. This is nothing new considering that every ring-meromorphic function $f(z)$ is changed by the substitution

$$z = e^{\zeta}$$

into a function meromorphic and periodic of period $2\pi i$ on the open ζ-plane. The validity of Picard's theorem for $f(\exp \zeta) = F(\zeta)$ yields an immediate proof of the above statement. If we write $\zeta = x + iy$ it follows that $F(\zeta)$ repeats its values in every infinite strip parallel to the x-axis of width 2π. Hence the surface $\mathfrak{F}$ on which these functions are defined presents itself as a circular cylinder of radius 1 around the x-axis. The order of a function on $\mathfrak{F}$ is described by a pair of functions ${}^eT(x)$ and ${}^iT(x)$:

$$T = [{}^eT(X);\ {}^iT(x)].$$

The first one is defined for $X > x_1$, the second one for $x < x_0$; $(x_1 > x_0)$, and they are subject to the condition

$$[{}^eT(X);\ {}^iT(x)] \sim [{}^eT(X) + cX;\ {}^iT(x) + cx].$$

Ring-meromorphic curves present themselves in this light as realizations in k-space of the cylindrical surface $\mathfrak{F}$ in abstracto.

Appendix

Suppose we are given a cell-division corresponding to a configuration $\{a\}$ of planes $a: \sum_0^k |a_i|^2 = 1$, in an $(k+1)$-dimensional vector space $\mathfrak{E}_{k+1}$: $\{\xi_0, \xi_1, \cdots, \xi_k\}$, of the sort described in Section 7. Concerning it we stated the following estimate: If b is a plane not passing through the center $\mathfrak{z}$ of a cell $\mathfrak{Z}$ then

$$\text{(1)} \qquad |(b\xi)| : \{\textstyle\sum_0^k |\xi_i|^2\}^{\frac{1}{2}} \geqq 2^{-k}\beta$$

for any point ξ of $\mathfrak{Z}$, where

$$\beta = |(b\mathfrak{z}^{(0)})| : \{\textstyle\sum_0^k |\mathfrak{z}_i^{(0)}|^2\}^{\frac{1}{2}},$$

$\mathfrak{z}^{(0)}$ being any point on the vertex $\mathfrak{z}$. In a conversation H. Weyl proposed to me the following proof of this inequality.

As a coordinate system in $\mathfrak{Z}$ we introduce its frame work $\{\zeta_0, \cdots, \zeta_k\}$. Let $(b\zeta) = 0$ be the equation of the plane b; then we have $\beta = |b_k| > 0$. For any point ζ for which $\sum_0^k |\zeta_i|^2 = 1$ (a restriction which does not impair the generality of our argument considering the conical shape of $\mathfrak{Z}$) we now write

$$|(b\zeta)| = s;$$

hence

$$|\zeta_0| \leqq s \qquad \text{on account of D 2).}$$

Consequently

$$(2) \qquad | b_1\zeta_1 + \cdots + b_k\zeta_k | \leqq s + | b_0\zeta_0 | \leqq s(1 + | b_0 |).$$

But from (7.2), applied to the space $\zeta_0 = 0$ of one dimension less, it follows that

$$(3) \qquad \begin{aligned} | b_1\zeta_1 + \cdots + b_k\zeta_k |^2 &\geqq \delta_k^2(1 - | b_0 |^2)(1 - | \zeta_0 |^2) \\ &\geqq \delta_k^2(1 - | b_0 |^2)(1 - s^2), \qquad (\delta_k > 0). \end{aligned}$$

The factor $(1 - | \zeta_0 |^2)$ replaces $| \zeta_1 |^2 + \cdots + | \zeta_k |^2$ which it equals since $\sum_0^k | \zeta_i |^2 = 1$, while $(1 - | b_0 |^2)$ makes its appearance because (7.2) was derived under the assumption that $\sum_0^k | b_i |^2 = 1$; hence in the present case we have to multiply on the right by $| b_1 |^2 + \cdots + | b_k |^2 = 1 - | b_0 |^2$ in order to make the relation completely analogous. Combining (2) and (3) we obtain

$$(4) \qquad \frac{s^2}{1 - s^2} \geqq \delta_k^2 \frac{1 - | b_0 |}{1 + | b_0 |}.$$

Hence, writing $| b_0 | = \cos \alpha_k$, the right side of (4) will become $\delta_k \operatorname{tg} (\alpha_k/2)$ and the inequality states that (7.2) will be satisfied for any $\delta_{k+1} = \delta$ chosen such that

$$\frac{\delta_{k+1}^2}{1 - \delta_{k+1}^2} \leqq \delta_k^2 \operatorname{tg}^2 \frac{\alpha_k}{2}.$$

Proceeding in this fashion we put successively

$$(5) \qquad \begin{aligned} | b_0 | : \sqrt{(| b_0 |^2 + \cdots + | b_k |^2)} &= \cos \alpha_k \geqq 0, \\ | b_1 | : \sqrt{(| b_1 |^2 + \cdots + | b_k |^2)} &= \cos \alpha_{k-1} \geqq 0, \\ &\vdots \\ | b_{k-1} | : \sqrt{(| b_{k-1} |^2 + | b_k |^2)} &= \cos \alpha_1 \geqq 0, \end{aligned}$$

counting the α's in inverse order. Solving with respect to the b's we have

$$(6) \qquad \begin{aligned} | b_0 | &= \cos \alpha_k, \\ | b_1 | &= \sin \alpha_k \cos \alpha_{k-1}, \\ &\vdots \\ | b_{k-1} | &= \sin \alpha_k \sin \alpha_{k-1} \cdots \cos \alpha_1, \\ \beta = | b_k | &= \sin \alpha_k \sin \alpha_{k-1} \cdots \sin \alpha_1. \end{aligned}$$

From the last equation under (6) it follows furthermore that none of the angles $\alpha_k, \alpha_{k-1}, \cdots, \alpha_1$ can be zero, and (5) completes the information to give

$$0 < \alpha_i \leqq \frac{\pi}{2}.$$

Therefore writing

$$0 < \operatorname{tg} \frac{\alpha_i}{2} = t_i \leqq 1$$

we have the recursion formulae

$$\delta_{i+1} = \delta_i t_i : \{1 + \delta_i^2 t_i^2\}^{\frac{1}{2}}$$

with $\delta_1 = 1$, $\delta_{k+1} = \delta$. Furthermore $\delta_i \leqq 1$; hence

$$\frac{\delta_{i+1}}{\delta_i} \geqq \frac{t_i}{\sqrt{(1 + t_i^2)}} \geqq \frac{t_i}{1 + t_i^2} = \frac{1}{2} \sin \alpha_i$$

and if we finally form the product we obtain

$$\prod_{i=1}^{k} \frac{\delta_{i+1}}{\delta_i} = \frac{\delta_{k+1}}{\delta_1} = \delta \geqq |b_k| \cdot 2^{-k} = 2^{-k}\beta,$$

which together with (7.2) completes the proof of (1).

URBANA, ILLINOIS.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

ON ADDING RELATIONS TO HOMOTOPY GROUPS

BY J. H. C. WHITEHEAD

1. Let X be an arcwise connected topological space, and let $\pi_r(X)$ ($r = 1, 2, \cdots$) be the r^{th} homotopy group[1] of X, written with multiplication if $r = 1$ and addition if $r > 1$. Let $f_i(S^{n-1}) \subset X$ ($i = 1, \cdots, k; n \geqq 2$) be maps in X of an $(n - 1)$-sphere S^{n-1}, let $\mathfrak{E}_i^n$ be a non-singular (open) cell bounded by $f_i(S^{n-1})$, as described below, and let

$$X^* = X + \mathfrak{E}_1^n + \cdots + \mathfrak{E}_k^n \qquad (\mathfrak{E}_i^n \cdot \mathfrak{E}_j^n = 0 \text{ if } i \neq j).$$

In a recent paper[2] I described in algebraical terms the relation between $\pi_{n-1}(X)$ and $\pi_{n-1}(X^*)$, and also the relation between $\pi_n(X)$ and $\pi_n(X^*)$ in case each of $f_i(S^{n-1})$ is homotopic to a point. Here we study the relation between $\pi_n(X)$ and $\pi_n(X^*)$ when the maps $f_i(S^{n-1})$ are arbitrary. There is a considerable difference between the cases $n = 2$ and $n > 2$. In case $n > 2$ the relation between $\pi_n(X)$ and $\pi_n(X^*)$ is expressed in terms of a product $\alpha \cdot \beta \in \pi_{m+n-1}(X)$, where $\alpha \in \pi_m(X)$, $\beta \in \pi_n(X)$. The case $n = 2$ is, in many ways, the more interesting of the two. Among other things a method is found for calculating[3] $\pi_2(K)$ algebraically, where K is any simplicial complex. Of course K. Reidemeister's[4] theory of homology in $\tilde{K}$, the universal covering complex of K, together with a theorem due to W. Hurewicz[5] lead to a theoretical definition of $\pi_2(K)$, which may be stated in purely algebraic terms. But since there is no general algorithm for deciding whether or no given elements $\rho_1, \cdots, \rho_n$ in the group ring, $\mathfrak{R}$, of K, satisfy given equations

$$\sum\nolimits_j \eta_{\lambda j}\rho_j = 0 \qquad (\eta_{\lambda j} \in \mathfrak{R})$$

this does not lead to a method of calculating $\pi_2(K)$. In fact the problem of calculating the algebraic structure of $\pi_2(K)$ by this method is equivalent to the problem of calculating $\pi_1(K)$ effectively, or of defining $\tilde{K}$ constructively. In order to calculate $\pi_2(K)$ itself one would also need a construction for a deforma-

[1] W. Hurewicz, Kon. Wetensch. Amsterdam, 38 (1935), 112-9; 521-8; 39 (1936), 117-25; 215-24.

[2] Proc. L. M. S., 45 (1939), 243-327, §6. This paper will be referred to as S.S. The argument given in S.S. obviously applies, with minor alterations, when K is any arcwise connected topological space.

[3] I.e., given a constructive definition of K, one can enumerate (constructively) a set of generators and relations for $\pi_2(K)$. Moreover, given a map $f(S^2) \subset K$, with a specified base point, one can express the corresponding element of $\pi_2(K)$ as a product of the generators, and conversely. We shall say that a group G has been calculated effectively when it has been calculated, and when a finite algorithm has been provided for deciding whether or no two given products of the generators represent the same element of G.

[4] See, among papers, K. Reidemeister, Abh. math. Sem. Hamb. 10 (1934), 211-5.

[5] Loc. cit. (Paper II), p. 522.

tion cell, bounded by a given circuit in $\tilde{K}$. It should be said that there is no theoretical obstacle to calculating $\pi_r(K)$, for any $r \geqq 1$, by constructions which are similar to those in a combinatorial definition of[6] $\pi_1(K)$. Thus the value of §6, below, is technical, rather than theoretical, in that it brings new algebraic machinery to bear on the study of $\pi_2(K)$.

We shall always use S^n to denote an oriented n-sphere, and E^n to denote an oriented n-element. The corresponding unoriented spaces will be denoted by $| S^n |$ and $| E^n |$. Let $f(\dot{E}^n) \subset X$ be a given map, where $\dot{E}^n$ is the boundary of E^n and E^n has no point in common with X, and let $\mathfrak{E}^n$ be the interior of E^n. By $X + \mathfrak{E}^n$ we shall mean the space consisting of the topological space X and the topological space $\mathfrak{E}^n$, related by the following condition. If $p_1, p_2, \cdots \subset \mathfrak{E}^n$ is an infinite set of points whose limit points all lie in the closed set $f^{-1}(q) \subset \dot{E}^n$, where q is any point in $f(\dot{E}^n) \subset X$, then the sequence $p_1, p_2, \cdots \subset X + \mathfrak{E}^n$ converges to q. Subject to this condition we describe $\mathfrak{E}^n$ as a non-singular cell bounded by $f(\dot{E}^n)$. When describing a geometrical construction we shall use the term *accidental intersection* to mean one which can be avoided without restricting the conditions of the problem in hand. For example, a point common to $\mathfrak{E}_i^n$ and $\mathfrak{E}_j^n$ $(i \neq j)$, in the above space X^*, would be an accidental intersection. Again, if we introduce a segment $s \subset M^n$, joining two given points p_1, p_2, where M^n is a connected, bounded manifold, then a point, other than p_1 or p_2, which in common to s and $\dot{M}^n$, would be an accidental intersection.

2. In this section we recall some elementary definitions in a convenient form. To define $\pi_n = \pi_n(X)$ we first choose a base point $x_0 \in X$. Then an element $\alpha \in \pi_n$ is given by a map $f(S^n, a) \subset X$, such that $f(a) = x_0$, where a is a specified base point in S^n. In general the same map $f(S^n)$ will represent a different element if another point $a' \in f^{-1}(x_0)$ is chosen as base point in S^n. The element $-\alpha$ is given by $f(-S^n, a)$, where $-S^n$ is S^n with the orientation reversed. Two maps $f_i(S_i^n, a_i)$ $(i = 1, 2)$, where $a_i \in S_i^n$ and $f_i(a_i) = x_0$, represent the same element of π_n if, and only if, $f_1^* = f_2\phi$, where $\phi(S_1^n) = S_2^n$ is a map of degree $+1$ such that $\phi(a_1) = a_2$, and $f_1^*(S_1^n, a_1)$ is homotopic in X, with $f_1(a_1)$ held fixed, to $f_1(S_1^n, a_1)$. Following Hurewicz we may also represent $\alpha \in \pi_n$ by a map $g(E^n) \subset X$ such that $g(\dot{E}^n) = x_0$, in which case we shall always take the map $g(\dot{E}^n) = x_0$ as the base point. Such a map will represent the same element as $f(S^n, a)$ if $g = f\phi$, where $\phi(E^n) = S^n$ is a map of degree $+1$, treating E^n as a cycle mod $\dot{E}^n$, such that $\phi(\dot{E}^n) = a$. Let $\alpha_i \in \pi_n$ $(i = 1, 2; n > 1)$ be given by $f_i(E_i^n) \subset X$, with $f_i(\dot{E}_i^n) = x_0$, and either

1. $| E_1^n | \cdot | E_2^n | = | \dot{E}_1^n | \cdot | \dot{E}_2^n | = | E^{n-1} |$ *and* $\dot{E}_1^n = E^{n-1} + \cdots, \dot{E}_2^n = -E^{n-1} + \cdots$, *or*
2. $| E_1^n | \cdot | E_2^n | = | \dot{E}_1^n | = | \dot{E}_2^n |$ *and* $\dot{E}_1^n = -\dot{E}_2^n$.

[6] Since neither can be calculated effectively, except in special cases, the only difference between the logical status of $\pi_1(K)$ and of $\pi_r(K)$ $(r > 1)$ is that $\pi_1(K)$, unlike $\pi_r(K)$, is always given by a finite system of generators and relations if K is finite.

In either case $\alpha_1 + \alpha_2 \in \pi_n$ is the element given by the map $f(E_1^n + E_2^n)$, with a point[7] on $\dot{E}_1^n$ as a base point in the second case, where $f = f_i$ in E_i^n.

Let $S_1^n = E_1^n - E^n$, $S_2^n = E^n - E_2^n$ and $S^n = E_1^n - E_2^n$, where $| E_1^n | \cdot | E_2^n | = | \dot{E}_i^n | = | \dot{E}^n |$. Let f be a given map of $K = | E_1^n | + | E_2^n | + | E^n |$ in X, with $f(a) = x_0$, where $a \in \dot{E}^n$, and let $\alpha_i \in \pi_n$ be the element given by $f(S_i^n, a)$.

LEMMA 1. *The map $f(S^n, a)$ represents the element $\alpha_1 + \alpha_2$.*

This follows at once from the fact that $f(K)$ is homotopic rel. a (i.e. holding $f(a)$ fixed) into a map $f_1(K)$ such that $f_1(E^n) = x_0$.

We recall from S.S., §§10, 11, that π_n $(n > 1)$ is a group with operators. As in S.S., §11, we shall consider the operators to be elements of the group ring $\mathfrak{R} = \mathfrak{R}(\pi_1)$, rather than elements of the ring $\mathfrak{R}_n(X, x_0)$, defined in[8] S.S. §10. If $\alpha \in \pi_n$, $\xi \in \pi_1$ the characteristic property of $\xi\alpha$ is that any of its representative maps can be transformed into a representative of α by a deformation in which the base point describes (positively) a circuit representing ξ. By an invariant sub-group $\sigma_n \subset \pi_n$, we shall mean one such that $\mathfrak{R}\sigma_n = \sigma_n$, that is to say $\rho\alpha \in \sigma_n$ if $\alpha \in \sigma_n$, $\rho \in \mathfrak{R}$. This is a natural generalization of the ordinary definition in case $n = 1$. We shall describe the groups defined on pp. 281 and 283 of S.S., which we shall now denote by $\mathfrak{R}(\alpha_1, \cdots, \alpha_k)$ and $\mathfrak{R}(f_1, \cdots, f_k)$ instead of $r(\alpha_1, \cdots, \alpha_k)$ and $r(f_1, \cdots, f_k)$, as the invariant sub-groups *generated* by $\alpha_1, \cdots, \alpha_k$ and by $f_1, \cdots, f_k$. This definition obviously applies to infinite sets of elements α_i or maps f_i, and in the same way we may define the invariant sub-group generated by a set of elements together with a set of maps.

Let an element $\alpha \in \pi_n$ $(n > 1)$ be given by a map $f(S^n, a) \subset X$ and let $f(a) = f(b) = x_0$. Let $s \subset S^n$ be an oriented segment beginning at b and joining it to a, and let ξ be the element of π_1, with base point x_0, which is represented by the circuit $f(s)$. Then it may be verified that $f(S^n, b)$ represents the element $\xi\alpha$. For this purpose S. Eilenberg's definition[9] of $\xi\alpha$ is particularly convenient.

3. Let $\alpha \in \pi_m = \pi_m(X)$ and $\beta \in \pi_n = \pi_n(X)$ $(m, n \geqq 1)$ be given elements represented by maps $f_0(E^m) \subset X$ and $g_0(E^n) \subset X$, such that $f_0(\dot{E}^m) = g_0(\dot{E}^n) = x_0$, where x_0 is to be taken as the base point for all the homotopy groups $\pi_r(X)$. We denote by $f_0 \cdot g_0$ the map

$$F_0(E^m \times E^n)^{\cdot} = F_0(\dot{E}^m \times E^n + (-1)^m E^m \times \dot{E}^n) \subset X,$$

given by

$$\begin{aligned} F_0(p \times q) &= g_0(q) \quad \text{if} \quad p \in \dot{E}^m, q \in E^n \\ &= f_0(p) \quad \text{if} \quad p \in E^m, q \in \dot{E}^n. \end{aligned}$$

[7] Since $n > 1$, $\dot{E}_1^n$ is connected and it is immaterial which point on $\dot{E}_1^n$ is taken as base point (see the concluding remark in this section).

[8] The distinction is that, if $\rho_1\alpha = \rho_2\alpha$ for each $\alpha \in \pi_n$, then ρ_1 and ρ_2 are identical, by definition, if regarded as elements of $\mathfrak{R}_n(X, x_0)$, but they may be different elements of $\mathfrak{R}$.

[9] Fund. Math., 32 (1939), 167–75. Eilenberg defines $\xi\alpha$ in terms of the universal covering space of X.

We take a point $a \times b \in \dot{E}^m \times \dot{E}^n$ as base point on $(E^m \times E^n)^{\cdot}$, where a is an arbitrary point on $\dot{E}^m$ if $m > 1$, and the first point of $\dot{E}^m$ if $m = 1$, and similarly for $b \in \dot{E}^n$. Let f_1 and g_1 be the images of f_0 and g_0 in homotopic deformations f_t and g_t $(0 \leqq t \leqq 1)$, such that $f_t(\dot{E}^m) = g_t(\dot{E}^n) = x_0$ for each $t \in \langle 0, 1 \rangle$. Then $f_1 \cdot g_1 = F_1$ is the image of F_0 in the deformation F_t, given by

$$\begin{aligned} F_t(p \times q) &= g_t(q) \quad \text{if} \quad p \in \dot{E}^m,\ q \in E^n \\ &= f_t(p) \quad \text{if} \quad p \in E^m,\ q \in \dot{E}^n, \end{aligned}$$

throughout which $F_t(a \times b) = x_0$. Therefore the element of $\pi_{m+n-1}(X)$ determined by the map $f_0 \cdot g_0$ depends only on the elements $\alpha \in \pi_m$, $\beta \in \pi_n$. We shall denote it by $\alpha \cdot \beta$. If $m = n = 1$ it is clear that

$$\xi \cdot \eta = \xi \eta \xi^{-1} \eta^{-1} \qquad (\xi, \eta \subset \pi_1), \tag{3.1}$$

and if $m = 1$, $n > 1$, that

$$\xi \cdot \beta = (\xi - 1)\beta. \tag{3.2}$$

If $m + n \geqq 2$ we have

$$\beta \cdot \alpha = (-1)^{mn} \alpha \cdot \beta \tag{3.3}$$

since $E^n \times E^m = (-1)^{mn} E^m \times E^n$.

THEOREM 1. *If $n > 1$ the transformation $\beta \to \alpha \cdot \beta$ is a homomorphism of π_n in π_{m+n-1} for each $\alpha \in \pi_m$.*

Let $\beta = \beta_1 + \beta_2$, where $\beta_i \in \pi_n$ $(i = 1, 2)$. Let β_i be represented by a map $g_i(E_i^n) \subset X$, where

$$| E_1^n | \cdot | E_2^n | = | \dot{E}_1^n | \cdot | \dot{E}_2^n | = | E^{n-1} |,$$

$$\dot{E}_1^n = E^{n-1} + \cdots, \qquad \dot{E}_2^n = -E^{n-1} + \cdots,$$

and $g_i(\dot{E}_i^n) = x_0$. Then β is given by $g(E^n)$, where $E^n = E_1^n + E_2^n$ and $g = g_i$ in E_i^n, and since $n > 1$ we may take a point in $\dot{E}^m \times E^{n-1}$ as the base point on $(E^m \times E^n)^{\cdot}$. Then the two $(m + n - 1)$-spheres

$$S_i^{m+n-1} = \dot{E}^m \times E_i^n + (-1)^m E^m \times \dot{E}_i^n$$

meet in the $(m + n - 1)$-element $E^m \times E^{n-1}$, and

$$S_1^{m+n-1} + S_2^{m+n-1} = \dot{E}^m \times E^n + (-1)^m E^m \times \dot{E}^n.$$

Therefore $\alpha \cdot \beta_1 + \alpha \cdot \beta_2 = \alpha \cdot (\beta_1 + \beta_2)$, by lemma 1, and the theorem is established.

If $m > 1$, $n > 1$ it follows from theorem 1 and (3.3) that the function $\alpha \cdot \beta$ determines a group multiplication.[10] However it may happen that $\alpha \cdot \beta = 0$ even though $\alpha \neq 0$, $\beta \neq 0$, as it does when $m = 1$ and X is n-simple in the sense of Eilenberg.[11]

[10] Cf. P. Alexandroff and H. Hopf, *Topologie*, Berlin (1935), 589–90.

[11] Loc. cit.

The homomorphism $\beta \to \alpha \cdot \beta$ $(n > 1)$ is not, in general, an operator homomorphism with respect to the operators in $\mathfrak{R}$. In fact if $m, n > 1$ it may be verified that

$$\xi(\alpha \cdot \beta) = \xi\alpha \cdot \xi\beta \qquad (\xi \in \pi_1).$$

If $m = 1, n > 1$ we have

$$\begin{aligned} \xi(\eta \cdot \beta) &= \xi(\eta - 1)\beta \\ &= (\xi\eta\xi^{-1} - 1)\xi\beta \\ &= \xi\eta\xi^{-1} \cdot \xi\beta \end{aligned}$$

for any $\eta \in \pi_1$, and if $m = n = 1$ the relation

$$\xi(\eta \cdot \eta')\xi^{-1} = \xi\eta\xi^{-1} \cdot \xi\eta'\xi^{-1} \qquad (\eta, \eta' \subset \pi_1)$$

follows from (3.1).

The product $\alpha \cdot \beta$ is one among many similar ways in which elements $\alpha \in \pi_m$ and $\beta \in \pi_n$ may be combined to form an element in π_{m+n-1}. For let α and β be given by $f(E^m) \subset X$ and $g(E^n) \subset X$, with $f(\dot{E}^m) = g(\dot{E}^n) = x_0$, where E^m and E^n are now the interior regions and boundaries of Euclidean metric spheres S^{m-1} and S^{n-1} in Euclidean spaces R^m and R^n. Let $p \to r_p$ $(p \in S^{m-1}, r_p \in G_n)$ be a continuous map of S^{m-1} in G_n, the group of rotations in R^n about the center of S^{n-1}, and let $q \to r_q$ $(q \in S^{n-1}, r_q \in G_m)$ be similarly defined. Then $\alpha \cdot \beta$ is a special case of the elements given by maps of the form $F(E^m \times E^n)^{\cdot} \subset X$, where

$$\begin{aligned} F(p \times q) &= g\{r_p(q)\} \quad \text{if} \quad p \in \dot{E}^m, q \in E^n \\ &= f\{r_q(p)\} \quad \text{if} \quad p \in E^m, q \in \dot{E}^n. \end{aligned}$$

Let $X = S^n$ $(n \geqq 2)$, let $m = 2$ let $\alpha = 0$, even when $n = 2$, and let $g(E^n) \subset S^n$, with $g(\dot{E}^n) = x_0$, be of degree 1. Then if $n = 2$ it may be verified, with the help of H. Hopf's invariant,[12] that the map $p \to r_p$ $(p \in S^1)$ determines an isomorphism between $\pi_1(G_n)$ and $\pi_{n+1}(S^n)$. The fact that $\pi_1(G_n)$ and $\pi_{n+1}(S^n)$ are both cyclic infinite if $n = 2$ and of order two[13] if $n > 2$ suggests that the same may be true if $n > 2$.

4. Let S_1^m and S_2^n $(1 < m \leqq n)$ be m- and n-spheres with a single common point b, and let $r = m + n - 1 > 1$. Let $\pi_{r1} = \pi_r(S_1^m)$, $\pi_{r2} = \pi_r(S_2^n)$ and let $\pi_m \cdot \pi_n$ be the sub-group of $\pi_r = \pi_r(S_1^m + S_2^n)$ which is generated by all elements of the form $\alpha \cdot \beta$, where $\alpha \in \pi_m(S_1^m)$, $\beta \in \pi_n(S_2^n)$ and b is the base point of π_{ri}, $\pi_m \cdot \pi_n$ and π_r.

THEOREM 2. *The group π_r is the direct sum $\pi_r(S_1^m) + \pi_r(S_2^n) + \pi_m \cdot \pi_n$. The group $\pi_m \cdot \pi_n$ is cyclic infinite.*

[12] Math. Ann., 104 (1931), 637–65.

[13] L. Pontrjagin, Oslo Congress (1936): H. Freudenthal, Compositio Math., 5 (1938), 299–314.

Let[14] $f(S^r) \subset S_1^m + S_2^n$ be a map representing a given element $\alpha^* \epsilon \pi_r$. We take S_1^m, S_2^n and S^r to be recti-linear sub-divisions of the boundaries of recti-linear simplexes, and we assume that the map f is simplicial. Let x be an inner point of an n-simplex in S_2^n. Then $f^{-1}(x)$ is a polyhedron whose cells may be oriented to form an $(m-1)$-cycle $Z^{m-1} \subset S^r$. Let $Z^m \subset S^r$ be a chain bounded by Z^{m-1} and let $\gamma_1(f)$ be the degree with which $f(Z^m)$ covers S_1^m. It follows from his original argument,[15] with trivial modifications, that $\gamma_1(f)$ is a 'Hopf invariant' of the element α^*, and it may therefore be written $\gamma_1(\alpha^*)$. Clearly $\alpha^* \to \gamma_1(\alpha^*)$ is a homomorphism of π_r in the additive group of integers.

We now prove the last part of the theorem. The groups $\pi_m(S_1^m)$ and $\pi_n(S_2^n)$ are cyclic infinite, generated by elements α and β, say. Therefore any element in $\pi_m \cdot \pi_n$ is of the form $k\alpha \cdot l\beta = kl(\alpha \cdot \beta)$, by theorem 1, where k and l are integers. Therefore $\pi_m \cdot \pi_n$ is a cyclic group generated by $\alpha \cdot \beta$. Obviously

$$\gamma_1(k\alpha \cdot l\beta) = \pm kl,$$

whence $k\alpha \cdot l\beta = 0$ implies $k = 0$ or $l = 0$. Taking $l = 1$ it follows that $k(\alpha \cdot \beta) \neq 0$ if $k \neq 0$, or that $\pi_m \cdot \pi_n$ is infinite.

We have now to prove that $\pi_r = \pi_{r1} + \pi_{r2} + \pi_m \cdot \pi_n$. Let $h_1(S_1^m + S_2^n) = S_1^m$ be the map given by

$$h_1(p) = p \quad \text{if} \quad p \epsilon S_1^m$$
$$= b \quad \text{if} \quad p \epsilon S_2^n.$$

Let $\phi_1(\pi_r) = \pi_{r1}$ be the homomorphism of π_r in which each element given by a map $f(S^r) \subset S_1^m + S_2^n$ is transformed into the element given by the map $h_1 f(S^r)$. Clearly $\phi_1(\pi_{r2}) = 0$ and $\phi_1(\alpha) = \alpha$ if $\alpha \epsilon \pi_{r1}$. Since $\alpha \cdot 0 = 0$, according to theorem 1, we have $\phi_1(\pi_m \cdot \pi_n) = 0$. If $\alpha_1 + \alpha_2 + \alpha \cdot \beta = 0$, where $\alpha_i \epsilon \pi_{ri}$, $\alpha \epsilon \pi_m(S_1^m)$, $\beta \epsilon \pi_n(S_2^n)$, it follows that

$$\alpha_1 = \phi_1(\alpha_1 + \alpha_2 + \alpha \cdot \beta) = 0.$$

Similarly $\alpha_2 = 0$ and hence $\alpha \cdot \beta = 0$. Therefore the group consisting of all elements of the form $\alpha_1 + \alpha_2 + \alpha \cdot \beta$ is the direct sum $\pi_{r1} + \pi_{r2} + \pi_m \cdot \pi_n$, and it remains to show that each element in π_r is of this form.

Let $\gamma_1(\alpha^*) = k$, where α^* is a given element in π_r, and, replacing α by $-\alpha$ if necessary, let $\gamma_1(\alpha \cdot \beta) = 1$, where $\alpha \cdot \beta$ generates $\pi_m \cdot \pi_n$. Then $\gamma_1(\alpha_0) = 0$, where $\alpha_0 = \alpha^* - k\alpha \cdot \beta$. Let $f(S^r) \subset S_1^m + S_2^n$ be a map representing α_0 and, as before, let $\dot{Z}^m = Z^{m-1}$, where $| Z^{m-1} | = f^{-1}(x)$. If $n > m$ it follows from a fundamental theorem due to Hopf[16] that $f(Z^m)$ can be deformed into a point, namely $f(Z^{m-1}) = x$, with $f(Z^{m-1})$ held fixed. Therefore an argument used by H. Freudenthal[17] shows that α_0 may be represented by a simplicial map f_1 such that $f_1^{-1}(x)$ is a single point p. Let $E^n \subset S_2^n$ and $E^r \subset S^r$ be the simplicial

[14] Since $\pi_1(S_1^m + S_2^n) = 1$ we need not specify which point in $f^{-1}(b)$ is to be taken as the base point in S^r.

[15] H. Hopf, loc. cit.: also Fund. Math., 25 (1935), 427–40.

[16] Comment. math. helv., 5 (1933), 39–54.

[17] Loc. cit., pp. 309–11.

neighborhoods of x and p, and let $\psi(S_2^n) = S_2^n$ be a map of degree unity such that $\psi(S_2^n - E^n) = b$, $\psi(E^n) = S_2^n$. We extend ψ throughout $S_1^m + S_2^n$ by writing $\psi(\eta) = y$ if $y \in S_1^m$. Clearly $\psi f_1(S^r)$ is homotopic to f_1, with the base-point, in $f_1^{-1}(b)$, held fixed. Therefore α_0 is also represented by the map ψf_1. But $\psi f_1(E^r) \subset S_2^n$, $\psi f_1(E_1^r) \subset S_1^m$ and $\psi f_1(S^{r-1}) = b$, where E_1^r is the closure of $S^r - E^r$ and $| S^{r-1} | = | \dot{E}^r | = | \dot{E}_1^r |$. Therefore $\alpha_0 = \alpha_1 + \alpha_2$ and $\alpha^* = \alpha_1 + \alpha_2 + k\alpha\cdot\beta$, where $\alpha_1 \in \pi_{r1}$ is represented by $\psi f_1(E_1^r) \subset S_1^m$, and $\alpha_2 \in \pi_{r2}$ by $\psi f_1(E^r) \subset S_2^n$. Therefore the theorem is established in case $m < n$.

Finally let $m = n$. With the same notation as before, let $\alpha^* = \alpha_0 + k\alpha\cdot\beta$, where α^* is a given element in π, and $\gamma_1(\alpha_0) = 0$, and let $\gamma_2(\alpha_0)$ be the degree with which $f(Z^m)$ covers S_2^n. Clearly $\gamma_1\{\phi_2(\alpha_0)\} = 0$, $\gamma_2\{\phi_2(\alpha_0)\} = \gamma_2(\alpha_0)$, where ϕ_2 is the homomorphism induced by the map h_2, given by $h_2(S_1^m) = b$, $h_2(p) = p$ if $p \in S_2^n$. Therefore $\gamma_1(\alpha_0^*) = \gamma_2(\alpha_0^*) = 0$, where $\alpha_0^* = \alpha_0 - \phi_2(\alpha_0)$, and it follows as when $n > m$ that $\alpha_0^* \in \pi_{r1} + \pi_{r2}$. But $\phi_2(\alpha_0) \in \pi_{r2}$ and $\alpha^* - \alpha_0 \in \pi_m\cdot\pi_n$. Therefore $\alpha^* = \alpha_0^* + (\alpha_0 - \alpha_0^*) + (\alpha^* - \alpha_0) \in \pi_{r1} + \pi_{r2} + \pi_m\cdot\pi_n$, and the proof is complete.

5. Let $f_i(S_i^{n-1}) \subset X_0$ $(i = 1, \cdots, k)$ be given maps of oriented $(n-1)$-spheres $S_1^{n-1}, \cdots, S_k^{n-1}$ in an arcwise connected space X_0. Let

$$X^* = X_0 + \mathfrak{E}_1^n + \cdots + \mathfrak{E}_k^n, \tag{5.1}$$

where $\mathfrak{E}_i^n$ is a non-singular cell bounded by $f_i(S_i^{n-1})$ and there are no accidental intersections. Let a_i^n be an open n-simplex, oriented in agreement with $\mathfrak{E}_i^n$, such that $A_i^n \subset \mathfrak{E}_i^n$, where $A_i^n = \bar{a}_i^n$, the closure of a_i^n. Then X_0 is a retract by deformation of

$$X = X_0 + \sum_{i=1}^{k} (\mathfrak{E}_i^n - a_i^n),$$

and it follows that the homotopy groups $\pi_r(X_0)$ and $\pi_r(X)$ are the same, likewise the relations between $\pi_r(X_0)$ and $\pi_r(X^*)$ and between $\pi_r(X)$ and $\pi_r(X^*)$. Also the identical map of $\dot{A}_i^n$ on itself is homotopic in X to $f_i(S_i^{n-1})$. Therefore we may replace X_0 by X and $f_i(S_i^{n-1})$ by the identical map of $\dot{A}_i^n$ on itself. Since $\mathfrak{E}_i^n$ may be triangulated and $A_i^n \subset \mathfrak{E}_i^n$, we may assume, after a suitable deformation, that any map $f(K) \subset X^*$, of a simplicial complex K, is simplicial in $f^{-1}(A_1^n + \cdots + A_k^n)$. From here to the last paragraph in this section we take $n > 2$.

Let $x_0 \in X$ be a base point for each $\pi_r = \pi_r(X)$ and let α_i be the element in π_{n-1} which is given by[18] $t_i + \dot{A}_i^n$, where t_i is an oriented segment starting at x_0 and joining it to a point $x_i \in \dot{A}_i^n$. Let $\mathfrak{M}$ be a modulus with independent basis elements $e_1, \cdots, e_k$ and coefficients in $\mathfrak{R} = \mathfrak{R}(\pi_1)$. Then the transformation ϕ, given by

$$\phi(\rho_1 e_1 + \cdots + \rho_k e_k) = \rho_1\alpha_1 + \cdots + \rho_k\alpha_k \qquad (\rho_i \in \mathfrak{R}) \tag{5.2}$$

[18] Cf. S.S., p. 279.

is a homomorphism of $\mathfrak{M}$ on $\mathfrak{R}(\alpha_1, \cdots, \alpha_k)$. Clearly $\mathfrak{R}\mathfrak{M}_0 = \mathfrak{M}_0$, where $\mathfrak{M}_0 = \phi^{-1}(0)$ is the kernel of ϕ.

Let $f(S^n, p_0) \subset X^*$, with $f(p_0) = x_0$, be given. After a suitable deformation, we assume that $f^{-1}(A_i^n)$ is a set of oriented n-simplexes $\epsilon_{i1}A_{i1}^n, \cdots, \epsilon_{iq_i}A_{iq_i}^n$, where $\epsilon_{i\lambda} = \pm 1$ $(i = 1, \cdots, k)$, $A_{i\lambda}^n$ takes its orientation from S^n, and $f(A_{i\lambda}^n) = \epsilon_{i\lambda}A_i^n$, the map f being linear in $A_{i\lambda}^n$. After a suitable sub-division of S^n and a further deformation[19] of f we assume that no two of the simplexes $A_{i\lambda}^n$, $A_{j\mu}^n$ have a common point. Let $s_{i\lambda} \subset S^n$ be an oriented, polygonal segment, starting at p_0 and joining it, without accidental intersections, to the point $p_{i\lambda} \in A_{i\lambda}^n$, such that $f(p_{i\lambda}) = x_i$, and let $t_{i\lambda} = f(s_{i\lambda})$. Then $f(s_{i\lambda} + \dot{A}_{i\lambda}^n) = t_{i\lambda} + \epsilon_{i\lambda}\dot{A}_i^n$ and is homotopic to $(t_{i\lambda} - t_i) + (t_i + \epsilon_{i\lambda}\dot{A}_i^n)$. Therefore the corresponding element in π_{n-1} is $\epsilon_{i\lambda}\xi_{i\lambda}\alpha_i$, where $\xi_{i\lambda} \in \pi_1$ is given by the circuit $t_{i\lambda} - t_i$. Then

$$\sum_{i,\lambda} \epsilon_{i\lambda}\xi_{i\lambda}\alpha_i = 0 \tag{5.3}$$

since $\sum_{i,\lambda}(t_{i\lambda} + \epsilon_{i\lambda}\dot{A}_i^n) = f\{\sum_{i,\lambda}(s_{i\lambda} + \dot{A}_{i\lambda}^n)\}$, and the singular sphere $\sum_{i,\lambda}(s_{i\lambda} + \dot{A}_{i\lambda}^n)$ bounds the cell $S^n - \sum_{i,\lambda}(s_{i\lambda} + A_{i\lambda}^n)$. Therefore

$$\psi(f) = \sum_{i,\lambda} \epsilon_{i\lambda}\xi_{i\lambda}e_i \in \mathfrak{M}_0. \tag{5.4}$$

Thus to the map f corresponds an element $\psi(f) \in \mathfrak{M}_0$. If $f(S^n) \subset X$, the set of simplexes $f^{-1}(A_i^n)$ being empty, we set $\psi(f) = 0$.

Lemma 2. *The element $\psi(f) = \psi(\alpha^*)$ depends only on the element $\alpha^* \in \pi_n^* = \pi_n(X^*)$, which is given by f.*

In the definition of $\psi(f)$, and in proving the lemma, it is obvious that a map of the form $f(S^n) \subset X^*$ may at any stage be replaced by one of the form $f(E^n) \subset X^*$, with $f(\dot{E}^n) = x_0$. Therefore the lemma will follow if we can show that $\psi(f_1) = \psi(f_2)$, where $f_i(E_i^n)$ $(i = 1, 2)$ are two given maps representing the same element $\alpha^* \in \pi_n^*$, both of which are simplicial in $f^{-1}(A_1^n + \cdots + A_k^n)$. We assume as we obviously may, that E_1^n and E_2^n are two hemispheres of an n-sphere $S^n = E_1^n - E_2^n$, where $\dot{E}_1^n = \dot{E}_2^n$, with regard to orientation. Then the map $f(S^n, p_0) \subset X^*$, where $p_0 \in \dot{E}_1^n$ and $f = f_i$ in E_i^n, represents the element $\alpha^* - \alpha^* = 0$. Clearly $\psi(f) = \psi(f_1) - \psi(f_2)$. So we have to show that $\psi(f) = 0$. Let $S^n = \dot{E}^{n+1}$. Then the map f can be extended throughout E^{n+1}, and we assume that $f(E^{n+1})$ is simplicial[20] in $f^{-1}(A_1^n + \cdots + A_k^n)$.

[19] Cf. S.S., p. 282.

[20] For it is given that f is simplicial in $f^{-1}(A_1^n + \cdots + A_k^n) \subset S^n$, and, among the processes of sub-division and canonical displacement of vertices by which $f(E^{n+1})$ is made simplicial in $f^{-1}(A_1^n + \cdots + A_k^n)$, there is obviously one which leaves $\psi(f)$ unaltered. Alternatively we can appeal to the following theorem, which is easy to prove, though I have not seen it in print. Let K and L be simplicial complexes and let $f_0(K) \subset L$ be a map which is simplicial in some sub-complex $K^* \subset K$. Then there is a sub-division K_1, of K, which leaves K^* untouched, and a deformation $f_t(K) \subset L$ $(0 \leqq t \leqq 1)$, such that $f_t = f_0$ in K^* and f_1 is simplicial with respect to K_1.

If $f(S^n) \subset X$ it is trivial that $\psi(f) = 0$. Otherwise let $f(A^n_{i\lambda}) = \epsilon_{i\lambda} A^n_i$, where $A^n_{i\lambda} \subset \dot{E}^{n+1}$, let y_i be an inner point of A^n_i, and let s be the segment in $f^{-1}(y_i)$ which, starting at $A^n_{i\lambda}$, terminates at some $A^n_{i\mu} \subset \dot{E}^{n+1}$. Let C^{n+1} be the chain of oriented $(n + 1)$-simplexes in $f^{-1}(A^n_i)$ which contain points of s. Then $\dot{E}^{n+1} = \dot{C}^{n+1} + \cdots$ and also $\dot{E}^{n+1} = A^n_{i\lambda} + A^n_{i\mu} + \cdots$. Therefore $\dot{C}^{n+1} = A^n_{i\lambda} + A^n_{i\mu} + \cdots$. Moreover $f(\dot{C}^{n+1} - A^n_{i\lambda} - A^n_{i\mu}) \subset \dot{A}^n_i$. Since $f(C^{n+1}) = A^n_i$, whence $f(\dot{C}^{n+1})$ is algebraically zero, we have $f(A^n_{i\lambda}) = -f(A^n_{i\mu})$, or $\epsilon_{i\lambda} = -\epsilon_{i\mu}$. Also $p_{i\lambda} \epsilon \dot{A}^n_{i\lambda}$ is joined to $p_{i\mu} \epsilon \dot{A}^n_{i\mu}$ by a segment $s^* \subset \dot{C}^{n+1}$, such that $f(s^*) = x_i$. Since the circuit $s_{i\lambda} + s^* - s_{i\mu}$ bounds a cell in E^{n+1}, and since $f(s^*) = x_i$, the circuit $f(s_{i\lambda}) - f(s_{i\mu})$ bounds a cell in X^*, and hence in X, since $n > 2$. Therefore $\xi_{i\lambda} = \xi_{i\mu}$. On repeating this argument it follows that, for each $i = 1, \cdots, k$, the simplexes $A^n_{i\lambda} \subset \dot{E}^{n+1}$ occur in pairs $A^n_{i\lambda}$, $A^n_{i\mu_\lambda}$, such that $\epsilon_{i\lambda} = -\epsilon_{i\mu_\lambda}$ and $\xi_{i\lambda} = \xi_{i\mu_\lambda}$. Therefore the expression (5.4) can be reduced to zero by cancelling terms of the form $\epsilon_{i\lambda}(\xi_{i\lambda} + \xi_{i\mu_\lambda})e_i$, where $\xi_{i\mu_\lambda} = \xi_{i\lambda}$. Therefore $\psi(f) = 0$, and the lemma is established.

It follows from an argument in the proof of lemma 2 that $\alpha^* \to \psi(\alpha^*)$ is a homomorphism of π^*_n in $\mathfrak{M}_0$.

LEMMA 3. *ψ is a homomorphism on $\mathfrak{M}_0$. It is an operator homomorphism, meaning that $\psi(\rho\alpha^*) = \rho\psi(\alpha^*)$ for any $\rho \epsilon \mathfrak{R}$. Its kernel is the sub-group $\pi^0_n \subset \pi^*_n$, which consists of elements with representative maps in X.*

Let α, given by (5.3), be a given element in $\mathfrak{R}(\alpha_1, \cdots, \alpha_k)$ and let $s_{i\lambda} + A^n_{i\lambda} \subset S^n$ mean the same as before. Then α is represented by a map $f(\Sigma) \subset X$, where

$$\Sigma = \sum_{i,\lambda} (s_{i\lambda} + \dot{A}^n_{i\lambda}),$$

such that $f(\dot{A}^n_{i\lambda}) = \epsilon_{i\lambda}\dot{A}^n_i$, $f(p_0) = x_0$, $f(p_{i\lambda}) = x_i$, and the circuit $f(s_{i\lambda}) - t_i$ represents the element $\xi_{i\lambda} \epsilon \pi_1$. The singular sphere Σ may be regarded as the boundary of the cell $S^n - \sum_{i,\lambda} (s_{i\lambda} + A^n_{i\lambda})$. Let e, given by (5.4), be an arbitrary element of $\mathfrak{M}_0$. Then $\alpha = 0$ and the map $f(\Sigma)$ can therefore be extended to a map $f(S^n - \sum_{i,\lambda} a^n_{i\lambda}) \subset X$, where $a^n_{i\lambda}$ is the interior of $A^n_{i\lambda}$, and hence to a map $f(S^n) \subset X^*$. Then $\psi(\alpha^*) = e$, where $\alpha^* \epsilon \pi^*_n$ is the element given by $f(S^n, p_0)$, and it follows that ψ is a homomorphism of π^*_n on $\mathfrak{M}_0$.

Let $\alpha^* \epsilon \pi^*_n$ be represented by a map $f(E^n) \subset X^*$, with $f(\dot{E}^n) = x_0$, and let f be simplicial in $f^{-1}(A^n_1 + \cdots + A^n_k)$. Let $A^n_{i\lambda}$ mean the same as before, but now let $s_{i\lambda}$ be a segment which, without accidental intersections, joins a point $p'_{i\lambda} \epsilon \dot{E}^n$ to $p_{i\lambda}$, where the points $p'_{11}, p'_{12}, \cdots$ are distinct. Let E^n_0 be an n-element such that $| E^n | \cdot | E^n_0 | = | \dot{E}^n | = | \dot{E}^n_0 |$, let p'_0 be an inner point of E^n_0 and let $s'_{i\lambda} \subset E^n_0$ be an oriented segment which starts at p'_0 and joins it to $p'_{i\lambda}$. Let ξ be any element in π_1, and let ξ be given by a map $g(b'b) \subset X$, with $g(b') = g(b) = x_0$, where $b'b$ is a simple segment. Let $f_0(E^n_0) = b'b$ be a map such that $f_0(p'_0) = b'$, $f_0(\dot{E}^n) = b$. Then $\xi\alpha^* \epsilon \pi^*_n$ is given by the map $f^*(S^n, p'_0) \subset X^*$, where $S^n = E^n + E^n_0$, $f^* = f$ in E^n and $f^* = gf_0$ in E^n_0. The circuit $f^*(s'_{i\lambda} + s_{i\lambda}) - t_i = f^*(s'_{i\lambda}) + f(s_{i\lambda}) - t_i$ represents the element $\xi\xi_{i\lambda} \epsilon \pi_1$.

Therefore $\psi(\xi\alpha^*) = \xi\psi(\alpha^*)$ and $\psi(\rho\alpha^*) = \rho\psi(\alpha^*)$, for any $\rho \in \mathfrak{R}$, since ψ is a homomorphism.

Clearly $\psi(\pi_n^0) = 0$. Conversely, let $\alpha^* \in \psi^{-1}(0)$ be given by a map $f(S^n, p_0) \subset X^*$, which is simplicial in $f^{-1}(A_1^n + \cdots + A_k^n)$. Then, since $e_1, \cdots, e_k$ are linearly independent, the expression (5.4) can be reduced to zero by cancelling pairs of terms of the form $(\xi_{i\lambda} - \xi_{i\mu})e_i$, where $\xi_{i\lambda} = \xi_{i\mu}$. If $\xi_{i\lambda} = \xi_{i\mu}$ the circuit $f(-s_{i\lambda} + s_{i\mu}) = -f(s_{i\lambda}) + f(s_{i\mu})$ is homotopic to a point in X. Therefore the circuit $f(s)$ can be deformed into a point, where s is any segment joining $p_{i\lambda}$ to $p_{i\mu}$, without accidental intersections. Let E_1^n, $E_2^n \subset S^n$ be n-elements such that

$$A_{i\lambda}^n + s + A_{i\mu}^n \subset \mathfrak{S}_1^n, \qquad E_1^n \subset \mathfrak{S}_2^n$$

and E_2^n does not meet p_0 or $A_{j\nu}^n$ $(j, \nu \neq i, \lambda$ or $i, \mu)$, where $\mathfrak{S}_h^n$ is the interior of E_h^n $(h = 1, 2)$. Then it follows from a standard argument[21] that $f(S^n)$ is homotopic, rel. $(S^n - \mathfrak{S}_2^n)$, to a map $f_1(S^n)$, such that $f_1(E_1^n) = x_i$, $f_1(E_2^n - \mathfrak{S}_1^n) \subset X$. Therefore the lemma follows from induction on the number of simplexes in $f^{-1}(A_1^n + \cdots + A_k^n)$.

Let α be any element of $\pi_n = \pi_n(x)$, given by a map $f(S^n) \subset X$, and let $\psi(\alpha) \in \pi_n^0$ be the element of π_n^* which is given by the same map. Then $\alpha \to \psi(\alpha)$ is obviously a homomorphism of π_n on π_n^0. Let σ_n be the invariant sub-group of π_n which is generated by maps of the form $f(S^n) \subset \dot{A}_i^n$ $(i = 1, \cdots, k)$ together with all elements of the form $\alpha_i \cdot \beta$, where $\alpha_1, \cdots, \alpha_k$ mean the same as before and $\beta \in \pi_2(X)$.

LEMMA 4. *σ_n is the kernel of the homomorphism $\psi(\pi_n) = \pi_n^0$.*

Clearly $\psi(\sigma_n) = 0$, and we have to show that, given a map $f(E^{n+1}) \subset X^*$ with $f(\dot{E}^{n+1}) \subset X$, then the class of elements $\mathfrak{R}\{f(\dot{E}^{n+1})\} \subset \pi_n$ is contained in σ_n. Let $y_i \in a_i^n$ and let the given map $f(E^{n+1})$ be simplicial in $f^{-1}(A_1^n + \cdots + A_k^n)$. Since $f(\dot{E}^{n+1}) \subset X$ it follows that $f^{-1}(y_1 + \cdots + y_k)$ is a set of simple, non-intersecting, polygonal circuits inside E^{n+1}. I say that the circuits in $f^{-1}(y_1 + \cdots + y_k)$ bound a set of non-singular, non-intersecting 2-elements. This is certainly the case if $n + 1 > 4$ since singularities and intersections between 2-cells can then be eliminated by slight deformations. If $n + 1 = 4$ $(n + 1 \geqq 4)$ since $n > 2)$ let s be any circuit in $f^{-1}(y_i)$. Then $s = \dot{E}^2$, where $E^2 \subset E^{n+1}$ is a non-singular, polyhedral 2-element, which does not meet $f^{-1}(y_1 + \cdots + y_k) - s$. For a suitable sub-division of E^{n+1} may be represented as a rectilinear sub-division of a rectilinear $(n + 1)$-simplex, and we may take E^2 to be a star whose center is in general position relative to $f^{-1}(y_1 + \cdots + y_k)$. Let E_0^{n+1} be a regular neighborhood[22] of E^2 which does not meet $\dot{E}^{n+1}$ or any of the other circuits in $f^{-1}(y_1 + \cdots + y_k)$. Then $\dot{E}_0^{n+1}$ may be joined to $\dot{E}^{n+1}$ by an $(n + 1)$-element $E_1^{n+1} \subset E^{n+1} - f^{-1}(y_1 + \cdots + y_k)$, such that $| E_1^{n+1} | \cdot | E_0^{n+1} | = | \dot{E}_1^{n+1} | \cdot | \dot{E}_0^{n+1} | = | E_0^n |$ and $| E_1^{n+1} | \cdot | E^{n+1} | = | \dot{E}_1^{n+1} | \cdot | \dot{E}^{n+1} | = | E^n |$. The

[21] See, for example, Alexandroff and Hopf (loc. cit.), p. 503; or S. Lefschetz, Fund. Math., 27 (1936), 94–115 (pp. 99–100).

[22] S.S., p. 293.

closure of $|E^{n+1}| - |E_1^{n+1}|$ meets E_1^{n+1} in the closure of $|\dot{E}^{n+1}| - |E^n|$, which is an n-element with no internal simplex in $\dot{E}^{n+1}$. Therefore the closure of $|E^{n+1}| - |E_1^{n+1}|$ is an $(n + 1)$-element. Similarly the closure of $|E^{n+1}| - |E_1^{n+1}| - |E_0^{n+1}|$ is an $(n + 1)$-element and the assertion follows from induction on the number of circuits in $f^{-1}(y_1 + \cdots + y_k)$. In the preceding argument, which also applies when $n + 1 > 4$, we cannot be certain that $f(\dot{E}_0^{n+1}) \subset X$, only that $f(\dot{E}_0^{n+1}) \subset X^* - (y_1 + \cdots + y_k)$. But we may replace A_i^n by a simplex $B_i^n \subset A_i^n$, such that $y_i \in B_i^n - \dot{B}_i^n$ and $f(\dot{E}_0^{n+1}) \subset X^* - (B_1^n + \cdots + B_k^n)$. This does not alter π_n, σ_n or π_n^0, or the relations between them. Therefore the lemma will follow from induction on the number of circuits in $f^{-1}(y_1 + \cdots + y_k)$ if we can show that $\mathfrak{R}\{f(\dot{E}_0^{n+1})\} \subset \sigma_n$, where A_i^n is replaced by B_i^n in the definition of σ_n. However, rather than this, we shall simplify the notation by starting again with $f(E^{n+1}) \subset X^*$, $f(\dot{E}^{n+1}) \subset X$, assuming that $f^{-1}(y_1 + \cdots + y_k) = f^{-1}(y_1)$, say, is a single circuit $\dot{E}^2$, where $E^2 \subset E^{n+1}$. The map f is to be simplicial in $f^{-1}(A_1^n)$, while E^2 is a sub-complex of some rectilinear subdivision of E^{n+1}. Therefore if the above simplex B_1^n is sufficiently small, it follows from a straightforward geometrical argument that, after starting again, $f^{-1}(\dot{A}_1^n)$ cuts E^2 in a simple circuit bounding a 2-element $E_0^2 \subset E^2$.

Let

$$K = E_0^2 + f^{-1}(\dot{A}_1^n).$$

I say that the identical map of $\dot{E}^{n+1}$ on itself is homotopic in $E^{n+1} - f^{-1}(a_1^n)$ to a map $u_2(\dot{E}^{n+1}) \subset K$. For the part of $f^{-1}(y_1)$ which lies in any $(n + 1)$-simplex $A^{n+1} \subset f^{-1}(A_1^n)$ is a linear segment, whose end points are internal to the two n-simplexes in $\dot{A}^{n+1}$ which cover A_1^n, whence it is clear that $f^{-1}(A_1^n)$ is a retract by deformation of $f^{-1}(A_1^n) - f^{-1}(y_1)$. Let u be a regular neighborhood of K in $E^{n+1} - f^{-1}(a_1^n)$ and let $E_0^{n+1} \subset u + f^{-1}(a_1^n)$ be a regular neighborhood of E^2, which is internal to E^{n+1} (even if u meets $\dot{E}^{n+1}$), and which contains E^2 in its interior. Then E^{n+1} contracts into[23] E_0^{n+1}, and it follows that the identical map $u_0(\dot{E}^{n+1}) = \dot{E}^{n+1}$ is deformable in $E^{n+1} - f^{-1}(y_1)$ into a map $u_1(\dot{E}^{n+1}) = \dot{E}_0^{n+1} \subset u + \{f^{-1}(a_1^n) - f^{-1}(y_1)\}$. The complex K is a retract by deformation of u, and hence a retract by a deformation in which each point of K is held fixed,[24] and $f^{-1}(\dot{A}_1^n)$ is a retract by a similar deformation, relative to $f^{-1}(\dot{A}_1^n)$, of $f^{-1}(A_1^n) - f^{-1}(y_1)$. Therefore the map $u_1(\dot{E}^{n+1})$ can be deformed, in $E^{n+1} - f^{-1}(y_1)$, into a map $u_2(\dot{E}^{n+1}) \subset K$, by deformations of the two parts lying in $f^{-1}(A_1^n) - f^{-1}(y_1)$ and in u, with the common part, in $f^{-1}(\dot{A}_1^n)$, held fixed. Since $f^{-1}(\dot{A}_1^n)$ is a retract by deformation of $f^{-1}(A_1^n) - f^{-1}(y_1)$, it follows that $E^{n+1} - f^{-1}(a_1^n)$ is a retract by deformation of $E^{n+1} - f^{-1}(y_1)$. Therefore the deformation cylinder of the deformation $u_0 \to u_2$ may be deformed into $E^{n+1} - f^{-1}(a_1^n)$, and the assertion is justified. It follows that $f(\dot{E}^{n+1})$ is deformable, in X, into the map $g(\dot{E}^{n+1}) \subset X$, where $g = fu_2$.

[23] S.S., pp. 248, 258, 260 and Theorem 23, corollary 1 (p. 293).
[24] S.S., p. 273.

Since $n > 2$ the circuit $f(\dot{E}_0^2) \subset \dot{A}_1^n$ is deformable, in $\dot{A}_1^n$, into the point x_1. This deformation can be extended to a deformation, in $\dot{A}_1^n$, of $f\{f^{-1}(A_1^n)\}$, and hence to a deformation $f_t(K) \subset X$ $(0 \leqq t \leqq 1; f_0 = f)$, such that $f_1(\dot{E}_0^2) = x_1$, $f_t\{f^{-1}(\dot{A}_1^n)\} = \dot{A}_1^n$. Therefore $f(\dot{E}^{n+1})$ is homotopic in X to $g_1(\dot{E}^{n+1})$, where $g_t = f_t u_2$. Let S^2 be a 2-sphere which meets A_1^n in the single point x_1, and let $h(\mathfrak{E}_0^2) = S^2 - x_1$ be a homeomorphism of $\mathfrak{E}_0^2 = E_0^2 - \dot{E}_0^2$ on $S^2 - x_1$. Let $h^*(K) = S^2 + \dot{A}_1^n$ be the map given by

$$\begin{aligned} h^* &= h \quad \text{in} \quad \mathfrak{E}_0^2 \\ &= f_1 \quad \text{in} \quad f^{-1}(\dot{A}_1^n), \end{aligned}$$

and $f^*(S^2 + \dot{A}_1^n) \subset X$ the map given by

$$\begin{aligned} f^*(x) &= f_1 h^{-1}(x) \quad \text{if} \quad x \in S^2 - x_1 \\ &= x \qquad\qquad \text{if} \quad x \in \dot{A}_1^n. \end{aligned}$$

Then $f_1(K) = f^*h^*(K)$, and

$$g_1(\dot{E}^{n+1}) = f^*g^*(\dot{E}^{n+1}) \subset X,$$

where $g^*(\dot{E}^{n+1}) = h^*u_2(\dot{E}^{n+1}) \subset S^2 + \dot{A}_1^n$. If we take x_1 as the base point for $\pi_n(S^2 + \dot{A}_1^n)$ it follows from theorem 2 that the element in $\pi_n(S^2 + \dot{A}_1^n)$, which is given by $g^*(\dot{E}^{n+1})$, is of the form $\beta_1 + \beta_2 + \alpha\cdot\beta$, where $\beta_1 \in \pi_n(S^2)$, $\beta_2 \in \pi_n(\dot{A}_1^n)$, $\alpha \in \pi_{n-1}(\dot{A}_1^n)$ and $\beta \in \pi_2(S^2)$. Clearly $g^*(\dot{E}^{n+1})$ is homotopic to a point in $S^2 + A_1^n$. Therefore $\beta_1 + \beta_2 + \alpha\cdot\beta$ reduces to zero if we fill in the simplex A_1^n. Since this has the same algebraic effect as mapping A_1^n on x_1 it follows that

$$\beta_1 = \phi_1(\beta_1 + \beta_2 + \alpha\cdot\beta) = 0,$$

where ϕ_1 means the same as in theorem 2, with S_1^m replaced by S^2 and S^n by $\dot{A}_1^n$. Let us transfer the base point of $\pi_n(X)$ to x_1, which is permissible since σ_n is an invariant sub-group, and let $\phi^*\{\pi_n(S^2 + \dot{A}_1^n)\} \subset \pi_n(X)$ be the homomorphism induced by the map $f^*(S^2 + \dot{A}_1^n) \subset X$. Then it follows from the definition of $\alpha\cdot\beta$ that $\phi^*(\alpha\cdot\beta) = \phi^*(\alpha)\cdot\phi^*(\beta)$ and hence that the element given by $f^*g^*(\dot{E}^{n+1})$ belongs to σ_n. This completes the proof.

Collecting these results we have the theorem:

THEOREM 3. *The residue group* $\pi_n^* - \pi_n^0$ *is isomorphic to* $\mathfrak{M}_0$, *and* $\pi_n - \sigma_n$ *is isomorphic to* π_n^0. *The homomorphism* $\psi(\pi_n^*) = \mathfrak{M}_0$ *determines an isomorphism* $\psi(\pi_n^* - \pi_n^0) = \mathfrak{M}_0$ *and* $\psi(\pi_n) = \pi_n^0$ *determines an isomorphism* $\psi(\pi_n - \sigma_n) \doteq \pi_n^0$.

Notice that lemma 2 is valid, for obvious reasons, when $n \geqq 2$, provided we take $\pi_1 = \pi_1(X^*)$, which coincides with $\pi_1(X)$ if $n > 2$. Let $\tilde{X}^*$ be the universal covering space of X^* and let $\tilde{X}$ be the part of $\tilde{X}^*$ which covers X. By an adaption of Reidemeister's theory, referred to in §1 above, we may interpret m as the group of relative n-cycles in $\tilde{X}^*$ (mod $\tilde{X}$). Now let $n = 2$ and let m_0 be the sub-group of m, which consists of the absolute (singular) 2-cycles in $\tilde{X}^*$. Then $\psi(\alpha^*) \subset m_0$, and ψ is a homomorphism of π_2^* in m_0. If X and X^* are

simplicial complexes it follows from Hurewicz' theorem, referred to in §1, that $\psi(\pi_2^*) = m_0$.

6. Let Γ be a given group, whose elements we denote by small greek letters, let H be an aggregate of individuals, which we denote by $a, b, a_i, b_j, \cdots$, and let $h(H) \subset \Gamma$ be a single-valued, but not necessarily $(1 - 1)$, transformation of H in Γ. We shall use h_Γ to stand for the group generated by all the pairs (a, ξ), which we denote by a_ξ, subject to the relations

$$a_{\xi\alpha} = a_\xi, \qquad a_\xi b_\eta = b_\zeta a_\xi, \tag{6.1}$$

for each $a, b \subset H$ and $\xi, \eta \subset \Gamma$, where $\alpha = h(a)$ and $\zeta = \xi\alpha\xi^{-1}\eta$. From the first of these, and induction on $|n|$ $(n = 0, \pm 1, \pm 2, \cdots)$, we have

$$a_{\xi\alpha^n} = a_\xi, \tag{6.2}$$

and it may be verified that the second implies

$$a_\xi^\delta b_\eta^\epsilon = b_\zeta^\epsilon a_\xi^\delta \qquad (\delta, \epsilon = \pm 1), \tag{6.3}$$

where $\zeta = \xi\alpha^\delta\xi^{-1}\eta$, with $\alpha = h(a)$. If $x = a_\xi^\delta \cdots b_\eta^\epsilon$ is any element in h_Γ, and $\tau \in \Gamma$, let

$$\theta_\tau(x) = a_{\tau\xi}^\delta \cdots b_{\tau\eta}^\epsilon .$$

Since $\tau(\xi\alpha) = (\tau\xi)\alpha$ and $(\tau\xi\alpha\xi^{-1}\tau^{-1})(\tau\eta) = \tau(\xi\alpha\xi^{-1}\eta)$, the transformation given by $x \to \theta_\tau(x)$, for each product of generators, leaves the system of relations invariant. It therefore determines a homomorphism of h_Γ in itself. Clearly $\theta_{\tau^{-1}}\theta_\tau(x) = \theta_\tau\theta_{\tau^{-1}}(x) = x$, whence θ_τ is an automorphism. Also $\theta_\tau\theta_{\tau'} = \theta_{\tau\tau'}$, whence $\tau \to \theta_\tau$ is a homomorphism of Γ in the group of automorphisms of h_Γ.

Let $\phi(a_\xi^\epsilon) = \xi\alpha^\epsilon\xi^{-1}$ for any $a_\xi \in h_\Gamma$, where $\alpha = h(a)$. Since $\xi\alpha\alpha\alpha^{-1}\xi^{-1} = \xi\alpha\xi^{-1}$ and

$$\begin{aligned} \xi\alpha\xi^{-1}\eta\beta\eta^{-1} &= \xi\alpha\xi^{-1}\eta\beta\eta^{-1}\xi\alpha^{-1}\xi^{-1}\xi\alpha\xi^{-1} \\ &= \zeta\beta\zeta^{-1}\xi\alpha\xi^{-1}, \end{aligned}$$

where $\beta = h(b)$ and $\zeta = \xi\alpha\xi^{-1}\eta$, the transformation given by

$$\phi(a_\xi^\delta \cdots b_\eta^\epsilon) = \phi(a_\xi^\delta) \cdots \phi(b_\eta^\epsilon)$$

is a homomorphism $\phi(h_\Gamma) \subset \Gamma$. Clearly $\phi(h_\Gamma)$ is the minimum invariant subgroup of Γ which contains $h(H)$. Notice also that $\phi\{\theta_\tau(x)\} = \tau\phi(x)\tau^{-1}$ for any $x \in h_\Gamma$, $\tau \in \Gamma$, whence $\theta_\tau\{\phi^{-1}(1)\} \subset \phi^{-1}(1)$. I say that $\phi^{-1}(1)$ is contained in the centrum of h_Γ. For let x and y be any elements in h_Γ, given by

$$x = a_{1\xi_1}^{\delta_1} \cdots a_{m\xi_m}^{\delta_m}, \qquad y = b_{1\eta_1}^{\epsilon_1} \cdots b_{n\eta_n}^{\epsilon_n}.$$

It follows from (6.3) and induction on $m + n$ that $xy = zx$, where

$$z = b_{1\zeta_1}^{\epsilon_1} \cdots b_{n\zeta_n}^{\epsilon_n}, \tag{6.4}$$

with[25] $\zeta_i = \phi(x)\eta_i$. Therefore $\phi(x) = 1$ implies $\zeta_i = \eta_i$, whence $z = y$. Let us now write the Abelian group $\phi^{-1}(1)$ with addition, and let us write $\theta_\tau(x) = \tau x$ for any $x \in \phi^{-1}(1)$. Then $\rho x \in \phi^{-1}(1)$ may be defined in the usual way, where ρ is any element in the integral group ring $\mathfrak{R}(\Gamma)$. Therefore $\phi^{-1}(1)$ is a group with operators in $\mathfrak{R}(\Gamma)$.

Returning to the point in §5 at which we required $n > 2$, we now take $n = 2$. Let $\alpha^* \in \pi_2^*$ be given by $f(S^2, p_0) \subset X^*$, which we assume to be simplicial in $f^{-1}(A_i^2)$. We shall simplify our notation by rewriting $s_{i\lambda}$, $p_{i\lambda}$, $\epsilon_{i\lambda}$ and $A_{i\lambda}^2 \subset S^2$ $(i = 1, \cdots, k;\ \lambda = 1, \cdots, q_i)$ as s_λ, p_λ, ϵ_λ and A_λ $(\lambda = 1, \cdots, q = q_1 + \cdots + q_k)$, where $f(A_\lambda) = \epsilon_\lambda A_{i_\lambda}^2$ and the segments $s_1, \cdots, s_q$ occur in this cyclic order round their common end point p_0. With a notation explained in S.S. p. 279, let $\Sigma = \Sigma_1 + \cdots + \Sigma_q$, where

$$\Sigma_\lambda = s_\lambda + \dot{A}_\lambda - s_\lambda ,$$

and Σ_λ does not meet Σ_μ except at p_0, and s_λ is non-singular and does not meet A_λ except at p_λ. The singular circuit

$$f(\Sigma) = f(\Sigma_1) + \cdots + f(\Sigma_q) \subset X,$$

in which $f(s_1)$ is described first, represents the element

$$\xi = \xi_1\alpha_{i_1}^{\epsilon_1}\xi_1^{-1} \cdots \xi_q\alpha_{i_q}^{\epsilon_q}\xi_q^{-1} \in \pi_1 , \tag{6.5}$$

where α_i means the same as in §5 and ξ_λ is the element given by the circuit $f(s_\lambda) - t_{i_\lambda}$. If h_{π_1} and $\phi(h_{\pi_1}) \subset \pi_1$ mean the same as before, with $\Gamma = \pi_1$, $H = (a_1, \cdots, a_k)$ and $h(a_i) = \alpha_i$, we have

$$\xi = \phi\{\psi(f, \Sigma)\},$$

where ξ is given by (6.5), and

$$\psi(f, \Sigma) = a_{i_1\xi_1}^{\epsilon_1} \cdots a_{i_q\xi_q}^{\epsilon_q}. \tag{6.6}$$

As in §5, the circuit $f(\Sigma)$ bounds the cell $f\left\{S^2 - \sum_{\lambda=1}^{q}(s_\lambda + A_\lambda)\right\}$. Therefore $\xi = 1$ and $\psi(f, \Sigma) \in \phi^{-1}(1)$. If $f(S^2) \subset X$ we set $\psi(f, \Sigma) = 0$.

Conversely, if x, given by the right hand side of (6.6), is an arbitrary element in h_{π_1}, we can construct $\Sigma \subset S^2$ and define a map $f(\Sigma) \subset X$, which represents the product $\xi = \phi(x)$, given by (6.5). If $\xi = 1$ the map $f(\Sigma)$ may be extended to a map $f(S^2) \subset X^*$ and we shall have $x = \psi(f, \Sigma)$. Therefore every element in $\phi^{-1}(1)$ is of the form $\psi(f, \Sigma)$ for a suitable choice of f and Σ.

Corresponding to lemmas 2 and 3 we have the theorem, in which π_n^0 $(n = 2)$ means the same as when $n > 2$:

THEOREM 4. *The element $\psi(\alpha^*) \in \phi^{-1}(1)$, given by* (6.6), *depends only on the element $\alpha^* \in \pi_2^*$, given by $f(S^2, p_0)$. The transformation $\alpha^* \to \psi(\alpha^*)$ is an operator homomorphism of π_2^* on $\phi^{-1}(1)$, and $\psi^{-1}(0)$ π_n^0*

[25] Notice the general form of the relations (6.3), namely $xyx^{-1} = \theta_{\phi(x)}(y)$.

In proving this we shall be concerned with $\phi^{-1}(1)$ as a sub-group of h_{π_1}, and shall therefore write it with multiplication. The first part of the theorem will follow from an argument used in lemma 2, when we have proved that $\psi(f, \Sigma) = 1$ if $\alpha^* = 0$. Therefore we assume that $S^2 = \dot{A}^3$, where A^3 is a rectilinear 3-simplex, and that there is a map $f(A^3) \subset X^*$. We also assume that f is simplicial in $f^{-1}(A_i^2)$. Let y_i be an inner point of A_i^2. Then, the trivial case $f(S^2) \subset X$ excepted, $f^{-1}(y_1 + \cdots + y_k) = L$, say, in a linkage which consists of non-singular, polygonal segments joining points on $\dot{A}^3$ and, possibly, circuits which are internal to A^3. We shall show that there is a homomorphism $F(G) \subset h_{\pi_1}$, where $G = \pi_1(A^3 - L)$, such that $\psi(f, \Sigma) = F(1) = 1$.

We take p_0 to be a vertex of A^3 and assume that the projection of L from p_0 on the opposite face is regular,[26] as the term is used in the theory of knots. Replacing $A_1^2, \cdots, A_k^2$ by smaller simplexes if necessary, we also assume that there are no accidental intersections in $f^{-1}(A_1^2 + \cdots + A_k^2)$; also that $T^2 = f^{-1}(\dot{A}_1^2 + \cdots + \dot{A}_k^2)$ cuts the cone swept out by the linear segment p_0p, as p varies over L, in a series of non-singular segments and circuits which are approximately parallel to the components of L. Let $L_1, \cdots, L_m$ be the segments of L which are 'completely visible' from p_0. That is to say the projection of L_ρ $(\rho = 1, \cdots, m)$ from p_0 does not pass under any segment of L, and each end point of L_ρ is either on $\dot{A}^3$, or is at a crossing of which L_ρ is a lower branch. Let $y_{i_\rho} = f(L_\rho)$, let $p \in L_\rho$, let p_ρ' be the point at which the rectilinear segment p_0p pierces T^2 and let p_ρ'' be a near-by point on T^2 such that $f(p_\rho'') = x_{i_\rho} \in \dot{A}_{i_\rho}^2$. Let c_ρ be a meridian circuit on T^2, beginning and ending at p_ρ'', which is oriented so that $f(c_\rho) = \dot{A}_{i_\rho}^2$, and let $l_\rho = p_0p_\rho' + p_\rho'p_\rho''$, where $p_0p_\rho' \subset p_0p$ and $p_\rho'p_\rho''$ is a segment on T^2 joining p_ρ' to p_ρ''. Then the group G is generated by $g_1, \cdots, g_m$, where g_ρ is given by the circuit $l_\rho + c_\rho + l_\rho$. We write

$$F(g_\rho) = a_{i_\rho \eta_\rho},$$

where $\eta_\rho \in \pi_1$ is given by the circuit $f(l_\rho) - t_{i_\rho}$. It follows from the same argument as when L is an ordinary knot or linkage that the relations determined by the crossings in the projection constitute a complete set.[27] Let

$$g_\rho^\epsilon g_\lambda = g_\mu g_\rho^\epsilon \tag{6.7}$$

be such a relation (see the diagram, in which $\epsilon = 1$).

The circuit $l_\rho + \epsilon c_\rho - l_\rho$ is obviously homotopic, rel. p_0, to a circuit of the form $l_\mu + p_\mu p_\lambda - l_\lambda$, where $p_\mu p_\lambda$ is a segment on T^2 which joins p_μ to p_λ. Since $f(p_\mu p_\lambda) \subset \dot{A}_{i_\mu}^2$ $(i_\mu = i_\lambda)$, the circuit $t_{i_\mu} + f(p_\mu p_\lambda) - t_{i_\lambda}$ represents an element of the form $\alpha_{i_\mu}^n$. Therefore $f(l_\mu + p_\mu p_\lambda - l_\lambda)$, which is homotopic, rel. p_0, to

$$\{f(l_\mu) - t_{i_\mu}\} + \{t_{i_\mu} + f(p_\mu p_\lambda) - t_{i_\lambda}\} - \{f(l_\lambda) - t_{i_\lambda}\},$$

---

[26] See, for example, K. Reidemeister, Knotentheorie, Berlin (1932), 5.

[27] See Knotentheorie, p. 54. This fact is, so to speak, half the content of the proof. For it follows from an argument given below that $\psi(f, \Sigma) = 1$ is equivalent to the relation corresponding to $\dot{A}^3$, when the latter is treated a single multiple point of the graph L.

represents the element $\eta_\mu \alpha^{\eta}_{i_\mu} \eta_\lambda^{-1}$. Similarly $f(l_\rho + \epsilon c_\rho - l_\rho)$ represents the element $\eta_\rho \alpha^{\epsilon}_{i_\rho} \eta_\rho^{-1}$. Therefore

$$(6.8) \qquad \eta_\mu \alpha^{\eta}_{i_\mu} = \eta_\rho \alpha^{\epsilon}_{i_\rho} \eta_\rho^{-1} \eta_\lambda .$$

From (6.3) we have

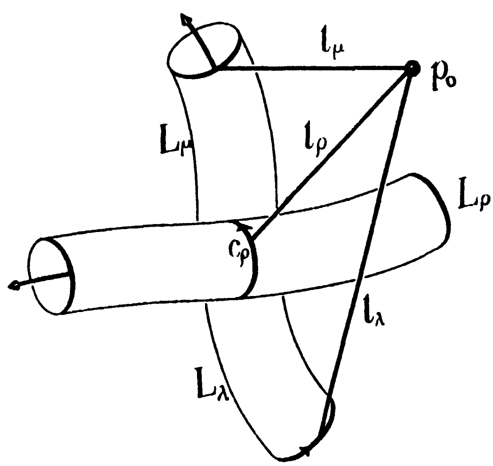
$$a^{\epsilon}_{i_\rho \eta_\rho} a_{i_\lambda \eta_\lambda} = a_{i_\lambda \zeta} a^{\epsilon}_{i_\rho \eta_\rho} ,$$

where $\zeta = \eta_\rho \alpha^{\epsilon}_{i_\rho} \eta_\rho^{-1} \eta_\lambda$, and it follows from (6.8) and (6.2) that

$$(6.9) \qquad a^{\epsilon}_{i_\rho \eta_\rho} a_{i_\lambda \eta_\lambda} = a_{i_\mu \eta_\mu} a^{\epsilon}_{i_\rho \eta_\rho} .$$

On comparing (6.7) with (6.9), we see that the transformation given by

$$F(g^{\delta}_{\lambda} \cdots g^{\epsilon}_{\mu}) = a^{\delta}_{i_\lambda \eta_\lambda} \cdots a^{\epsilon}_{i_\mu \eta_\mu}$$



is a homomorphism $F(G) \subset h_{\pi_1}$. Let $\phi_f(G) \subset \pi_1$ be the homomorphism of G determined by the map f and the base point $f(p_0) = x_0$. Then $\phi_f(g_\rho) = \eta_\rho \alpha_{i_\rho} \eta_\rho^{-1}$, whence $\phi_f = \phi F$.

Let c be a meridian circuit on T^2, oriented so that $f(c) = \dot{A}^2_i$ for some value of i, and let $l \subset A^3 - L$ be any segment which joins p_0 to a point $p \in f^{-1}(x_i) \cdot c$. Let $g \in G$ be the element given by $l + c - l$, and let $\xi \in \pi_1$ be the element given by $f(l) - t_i$. I say that

$$(6.10) \qquad F(g) = a_{i\xi} .$$

For let pp''_ρ be a segment on T^2 which joins p to some p''_ρ. Then g is also given by $(l + pp''_\rho) + c_\rho - (l + pp''_\rho)$. On the other hand

$$f(l + pp''_\rho) - t_i = f(l) + f(pp''_\rho) - t_i ,$$

which is homotopic, rel p_0, to

$$\{f(l) - t_i\} + \{t_i + f(pp''_\rho) - t_i\},$$

and since $f(pp_\rho'') \subset \dot{A}_i^2$ it follows that the element of π_1 which is given by $f(l + pp_\rho'') - t_i$ is of the form $\xi\alpha_i^n = \eta$, say, whence $a_{i\xi} = a_{i\eta}$, in consequence of (6.2). Therefore we may replace c by c and l by $l + pp_\rho''$ without altering either g of $a_{i\xi}$. To simplify the notation we shall start again, assuming that $c = c_p$, $p = p_\rho''$. Then $g = \bar{g}g_\rho\bar{g}^{-1}$, where $\bar{g}$ is the element given by $l - l_\rho$, and

$$F(g) = \bar{x}a_{i\eta_\rho}\bar{x}^{-1} \qquad (i = i_\rho),$$

where $\bar{x} = F(\bar{g})$. It follows from (6.4) that

$$F(g) = a_{i\xi},$$

where $\xi = \phi(\bar{g})\eta_\rho$. But $\phi(\bar{x}) = \phi F(\bar{g}) = \phi_f(\bar{g})$, and since $\bar{g}$ is given by $l - l_\rho$ the element $\phi(\bar{x})$ is given by $f(l) - f(l_\rho)$. Therefore $\xi = \phi(\bar{x})\eta_\rho$ is given by $f(l) - f(l_\rho) + f(l_\rho) - t_i$, or by $f(l) - t_i$, as stated.

It follows from the preceding paragraph that $F(g^*) = \psi(f, \Sigma)$, where $g^* \in G$ is given by the circuit Σ. But $g^* = 1$, since Σ bounds a cell in $\dot{A}^3$. Therefore $\psi(f, \Sigma) = 1$, and the first part of the proof is complete.

The fact that ψ is a homomorphism follows from the argument used to reduce the first part to the case where $\alpha^* = 0$. The argument used when $n > 2$ shows that ψ is an operator homomorphism.

It follows from the definition that $\psi(\pi_2^0) = 1$, the group $\phi^{-1}(1)$ still being multiplicative. Conversely let $\psi(\alpha^*) = 1$ and let α^* be given by a map $f(S^2) \subset X^*$ which is simplicial in $f^{-1}(A_1^2 + \cdots + A_k^2)$. Then the product (6.6), determined by the map f and some $\Sigma \subset S^2$, represents the identity in h_{π_1}. Therefore it can be transformed into a product of the form

$$x_1R_1^{\epsilon_1}x_1^{-1} \cdots x_nR^{\epsilon_n}x_n^{-1}, \tag{6.11}$$

where each R_λ is of the form

$$a_{i\xi\alpha_i}a_\xi^{-1} \qquad \text{or} \qquad a_{i\xi}a_{j\eta}a_{i\zeta}^{-1}a_{j\zeta}^{-1} \qquad (\zeta = \xi\alpha_i\xi^{-1}\eta),$$

by a sequence of operations which consist of cancelling, or of inserting, consecutive terms of the form $a_{i\xi}^{\epsilon}a_{i\xi}^{-\epsilon}$. Each cancelling operation can be copied by the geometrical process described in lemma 3, which can obviously be reversed so as to copy an insertion. Therefore we may assume in the first place that the product (6.6) is of the form (6.11). Moreover we may subject the factors $a_{i\xi}^{\epsilon}$ of (6.6) and, at the same time, the subscripts $1, \cdots, q$ in $\Sigma = \Sigma_1 + \cdots + \Sigma_q$, to any cyclic permutation. Therefore, after transferring x_1, from the beginning to the end of (6.11), and inserting $x_1x_1^{-1}$ between x_λ^{-1} and $x_{\lambda+1}$, for each $\lambda = 2, \cdots, n - 1$, we may further assume that $x_1 = 1$.

The geometrical process by which terms of the form xx^{-1} are removed involves an alteration of the map f. We now leave f as it is, but replace $s_1, \cdots, s_q$ by segments $s_1', \cdots, s_q'$, in such a way as to transform $\psi(f, \Sigma)$, given by (6.1), into

$$\psi(f, \Sigma') = Wx_2R_2^{\epsilon_2}x_2^{-1} \cdots x_nR_n^{\epsilon_n}x_n^{-1},$$

where W is of the form aa^{-1} or $aa^{-1}bb^{-1}$.

First let $R_1 = a_{i\xi\alpha_i}a_{i\xi}^{-1}$. Then $R_1^{-1} = a_{i\eta\alpha_i^{-1}}a_{i\eta}^{-1}$, where $\eta = \xi\alpha_i$, and replacing ξ by $\xi\alpha_i^{-1}$ if $\epsilon = -1$, we have $R_1^\epsilon = a_{i\xi\alpha_i^\epsilon}a_{i\xi}^{-1}$. Let

$$\Sigma' = s_1' + \dot{A}_1' - s_1' + \cdots + s_q' + \dot{A}_q' - s_q',$$

where $A_\lambda' = A_\lambda$ ($\lambda = 1, \cdots, q$) and

$$s_1' = s_1 - \epsilon\dot{A}_1 \quad \text{(S.D.)}, \qquad s_\lambda' = s_\lambda \qquad (\lambda = 2, \cdots, q),$$

in which (S.D.) means that the segment indicated is to be slightly deformed so as to eliminate accidental intersections. Then, remembering that $f(A_1) = A_i^2$, we see that $f(s_1') - t_i$ is homotopic, rel. p_0, to

$$f(s_1) - t_i + (t_i - \epsilon\dot{A}_i^2 - t_i).$$

Since $f(s_1) - t_i$ represents the element $\xi\alpha_i^\epsilon$, it follows that the circuit $f(s_1') - t_i$ represents the element $\xi\alpha_i^\epsilon\alpha_i^{-\epsilon} = \xi$. Therefore

$$\psi(f, \Sigma') = a_{i\xi}a_{i\xi}^{-1}x_2R_2^{\epsilon_2}x_2^{-1} \cdots x_nR_n^{\epsilon_n}x_n^{-1}. \tag{6.12}$$

Secondly let $R_1 = a_{i\xi}a_{j\eta}a_{i\xi}^{-1}a_{j\zeta}^{-1}$, where $\zeta = \xi\alpha_i\xi^{-1}\eta$. Then $R_1^{-1} = a_{j\zeta}a_{i\xi}a_{j\eta}^{-1}a_{i\xi}^{-1}$, and if $\epsilon_1 = -1$ we replace every factor $x_\lambda R_\lambda^{\epsilon_\lambda}x_\lambda^{-1}$ in (6.11) by $a_{j\zeta}^{-1}x_\lambda R_\lambda^{\epsilon_\lambda}x_\lambda^{-1}a_{j\zeta}$ and cancel the first two terms $a_{j\zeta}^{-1}a_{j\zeta}$. This operation can be copied geometrically, and the result is to replace R_1^{-1} by $a_{i\xi}a_{j\eta}^{-1}a_{i\xi}^{-1}a_{j\zeta}$. Therefore we may take $R_1^\epsilon = a_{i\xi}a_{j\eta}^\epsilon a_{i\xi}^{-1}a_{j\zeta}^{-\epsilon}$, Let

$$\Sigma' = s_1' + \dot{A}_1' - s_1' + \cdots + s_q' + \dot{A}_q' - s_q',$$

where

$$s_1' = s_1 + \dot{A}_1 - s_1 + s_2 \quad \text{(S.D.)}, \qquad s_2' = s_1, \qquad s_\lambda' = s_\lambda,$$

$$A_1' = A_2, \qquad A_1' = A_1, \qquad A_\lambda' = A_\lambda \qquad (\lambda = 3, \cdots, q).$$

Then $f(A_1') = A_j^2$, $f(A_2') = A_i^2$ and $f(s_1') - t_j$ is homotopic to

$$\{f(s_1) - t_i\} + (t_i + \dot{A}_i^2 - t_i) - \{f(s_1) - t_i\} + \{f(s_2) - t_j\},$$

and therefore represents the element $\xi\alpha_i\xi^{-1}\eta = \zeta$. The circuit $f(s_2') - t_i = f(s_1) - t_i$ represents the element ξ, and we have

$$\psi(f, \Sigma') = a_{j\zeta}^\epsilon a_{i\xi}a_{i\xi}^{-1}a_{j\zeta}^{-\epsilon}x_2R_2^{\epsilon_2}x_2^{-1} \cdots x_nR^{\epsilon_n}x_n^{-1}. \tag{6.13}$$

If $n = 1$ it follows from (6.12) or (6.13) that $f(S^2)$ is homotopic to a map in X. In general, the terms preceding x_2, in (6.12) or in (6.13), can be removed by cancelling, and it follows from induction on n that $\alpha^* \epsilon \pi_2^0$. Therefore π_2^0 is the kernel of ψ and the theorem is established.

Corollary. *If $\pi_2(X) = 0$, then $\pi_2(X^*)$ is isomorphic to $\phi^{-1}(1)$ under the transformation ψ.*

As an application of this corollary, let X^* be a 2-dimensional complex and let $\mathfrak{E}_1^n, \cdots, \mathfrak{E}_k^n$ be all the 2-cells in X^*, with the notation used at the beginning of §5. Then X_0 is the linear graph which consists of all the 1-cells in X^*. There-

fore $\pi_2(X) = 0$ and $\pi_1(X)$ is a free group. Therefore, given a product $a_\xi^\delta \cdots b_\eta^\epsilon$, of the generators of h_{π_1}, we can decide whether or no $\phi(a_\xi^\delta \cdots b_\eta^\epsilon) = 1$. Let F be the free group, which is freely generated by the generators of h_{π_1}, and let $\phi(x) \in \pi_1(X)$ have the obvious meaning if $x \in F$. Then the (multiplicative) group π_2^* can be calculated in the form:

Generators: *all the elements of* $\phi^{-1}(1) \subset F$,
Relations: *all relations of the form* $x = y$, *where* $x, y \subset \phi^{-1}(1)$ *are of the form* $x = x_1x_2$, $y = x_1Rx_2$, *in which* R *is of the form* $a_{\xi\alpha}a_\xi^{-1}$ or $a_\xi b_\eta a_\zeta^{-1} b_\xi^{-1}$ $(\alpha = \phi(a), \zeta = \xi\alpha\xi^{-1}\eta)$.

By this method we not only calculate π_2^* as an abstract group but, given a map $f(S^2, p_0) \subset X^*$, with $f(p_0) = x_0$, we can calculate $\psi(\alpha^*) \in \phi^{-1}(1)$, where α^* is the element represented by $f(S^2, p_0)$. Conversely, given $x \in \phi^{-1}(1)$ we can construct a map which represents $\psi^{-1}(x) \in \pi_2^*$, as in the preamble to theorem 4. Therefore, given any simplicial complex K, we can first calculate $\pi_2(K^2)$ and then $\pi_2(K^3) = \pi_2(K)$, by means of S. S., theorem 18, where K^n is the n-dimensional skeleton of K.

If X^* is a polyhedron the kernel of the homomorphism $\chi(\pi_2) = \pi_2^0$ can be expressed in terms of Reidemeister's[28] theory of homology with coefficients in $\mathfrak{R}$, or in the residue ring $\mathfrak{R} - \mathfrak{J}$, where $\mathfrak{J}$ is any two-sided ideal in $\mathfrak{R}$. For let $\tilde{X}^*$ be the universal covering space of X^*, and let $\tilde{X} = u^{-1}(X)$, where $u(\tilde{X}^*) = X^*$ is a regular covering of X^* by $\tilde{X}^*$. Let $\tilde{\tilde{X}}$ be the universal covering space of $\tilde{X}$, and hence of X, and let $\bar{u}(\tilde{\tilde{X}}) = \tilde{X}$ be a regular covering of $\tilde{X}$ by $\tilde{\tilde{X}}$. Then[29]

$$\pi_2(X) \simeq \pi_2(\tilde{\tilde{X}}) \simeq \beta_2(\tilde{\tilde{X}}), \qquad \pi_2(X^*) \simeq \pi_2(\tilde{X}^*) \simeq \beta_2(\tilde{X}^*),$$

where $\simeq$ denotes isomorphism and $\beta_2(P)$ is the second homology group of P, with integral coefficients. More precisely, $\pi_2(\tilde{\tilde{X}})$ is isomorphic to $\beta_2(\tilde{\tilde{X}})$ in the transformation under which $\alpha \in \pi_2$, given by $f(S^2) \subset \tilde{\tilde{X}}$, corresponds to the homology class containing the cycle $f(S^2)$. The homomorphism $\chi(\pi_2) = \pi_2^0$ is the one determined by the map $\bar{u}(\tilde{\tilde{X}}) \subset \tilde{X}^*$. It follows that the homology classes corresponding to the elements in $\chi^{-1}(0)$ are those which reduce to zero in consequence of the relations $\alpha_1 = \cdots = \alpha_k = 0$.

In case X is a finite, 2-dimensional complex one can also express $\chi^{-1}(0)$ as follows. Let $X = X_1 = K^1 + B_1^2 + \cdots + B_m^2$, $X^* = X + A_1^2 + \cdots + A_k^2$, and let $X_2 = K^1 + A_1^2 + \cdots + A_k^2$, where K^1 is a linear graph. Let us rewrite the group h_{π_1}, which is determined by X and X^*, as $h(X, X^*)$. Then $h(K^1, X^*)$ is generated by the generators $a_{i\xi}$ $(i = 1, \cdots, k)$ of $h(K^1, X_1)$, together with the generators $b_{j\eta}$ $(j = 1, \cdots, m)$ of $h(K^1, X_2)$, where $\xi, \eta \subset F = \pi_1(K^1)$. The relations for $h(K^1, X^*)$ consist of the relations $R_{\rho\lambda}$, for $h(K^1, X_\rho)$ $(\rho = 1, 2;$ $\lambda = 1, 2, \cdots)$, together with a system of relations $R_{12\lambda}$, which are of the form

$$a_{i\xi}b_{j\eta} = b_{j\zeta}a_{i\xi}.$$

[28] Loc. cit.
[29] Hurewicz (loc. cit.), paper II, p. 522.

Then $\chi^{-1}(0)$ is isomorphic to the sub-group of $\phi^{-1}(1) \subset h(K^1, X_1)$, whose elements reduce to 1 on the introduction of the new generators $b_{j\eta}$ and the additional relations $R_{2\lambda}$ and $R_{12\lambda}$. In particular $\pi_2 = \pi_2^0$ if every relation between the generators $a_{i\xi}$ is a consequence of the relations $R_{1\lambda}$.

By way of application consider the question: *Is any sub-complex of an aspherical,*[30] *2-dimensional complex itself aspherical?* Let $K^2 = K^1 + A_1^2 + \cdots + A_k^2 + B_1^2 + \cdots + B_m^2$ be any 2-dimensional complex, let $K_1^2 = K^1 + B_1^2 + \cdots + B_m^2$, $K_2^2 = K^1 + A_1^2 + \cdots + A_k^2$ and let us denote arbitrary elements of $h(K^1, K_1^2)$, of $h(K^1, K_2^2)$ and of $h(K^1, K^2)$ by x, by y and by z or z' respectively. We recall the general form of the relations for $h(K^1, K^2)$ which follows from (6.4), namely $zz'z^{-1} = g_z(z')$, where $g_z = \theta_{\phi(z)}$. Since $z \to \phi(z)$ and $\zeta \to \theta_\zeta$ ($\zeta \in \pi_1$) are homomorphisms, it follows that $z \to g_z$ is a homomorphism of $h(K^1, K^2)$ in its group of automorphisms. The group $h(K^1, K^2)$ is obtained from the free product $h(K^1, K_1^2) \circ h(K^1, K_2^2)$ by adjoining the relations $xyx^{-1} = g_x(y)$, $yxy^{-1} = g_y(x)$, for every $x \in h(K^1, K_1^2)$, $y \in h(K^1, K_2^2)$. Thus the above question is part of a wider question, which can be stated as follows. Let G_1 and G_2 be given groups, and to each $x \in G_1$, $y \in G_2$, let there correspond automorphisms $y' = f_x(y)$, $x' = g_y(x)$, of G_2 and G_1, such that the transformations $x \to f_x$ and $y \to g_y$ are homomorphisms of G_1 and G_2 in the automorphic groups of G_2 and G_1. Then the question is: *Under what conditions are G_1 and G_2, regarded as sub-groups of $G_1 \circ G_2$, unaffected by the additional relations $xyx^{-1} = f_x(y)$, $yxy^{-1} = g_y(x)$, in which x and y range over all the elements in G_1 and G_2?*

BALLIOL COLLEGE
OXFORD, ENGLAND.

[30] Hurewicz (loc. cit.), paper IV. See also J. H. C. Whitehead, Fund. Math., 32 (1939), 149–66.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

ON HUMBERT FUNCTIONS

BY R. S. VARMA

(Received December 12, 1939)

I. INTRODUCTION

The function $J_{m,n}(x)$ defined by the relation

$$J_{m,n}(x) = \frac{(x/3)^{m+n}}{\Gamma(m+1)\Gamma(n+1)}\, {}_0F_2\left(m+1, n+1; -\frac{x^3}{27}\right) \tag{1}$$

was first studied by P. Humbert[1] in the year 1930 and later on, among other results, he gave the following operational relations:[2]

$$x^{\frac{2m-n}{3}} J_{m,n}(3\sqrt[3]{x}) \risingdotseq \frac{1}{p^{\frac{2m-n}{2}}} J_n\left(-2\sqrt{\frac{1}{p}}\right) \tag{2}$$

and

$$x^{\frac{2n-m}{3}} J_{m,n}(3\sqrt[3]{x}) \risingdotseq \frac{1}{p^{\frac{2n-m}{2}}} J_m\left(-2\sqrt{\frac{1}{p}}\right). \tag{3}$$

The function $J_{m,n}(x)$ has been called by Humbert a Bessel function of the third kind, but in order to avoid confusion with the ordinary Bessel function of the third kind, I shall call it a Humbert function.

The object of this paper is to investigate[3] some properties of Humbert functions so far as the convergence of infinite series involving the functions are concerned. In general we can have the following two types of infinite series involving the functions:

$$\sum_{n=1}^{\infty} A_{m,n} J_{m,n}(x) \tag{A}$$

and

$$\sum_{n=1}^{\infty} A_{ln+k,n} J_{ln+k,n}(x). \tag{B}$$

---

[1] P. Humbert: Les fonctions de Bessel du troisieme ordre, Atti. Pont. Acad. della Scienza, Anno. LXXXIII (Sess. III del 16 Febbraio, 1930), 128–146.

[2] P. Humbert: Nouvelles remarques sur les fonctions de Bessel du troisieme ordre, ibid, Anno. LXXXVII (Sess. IV del 18 Marzo, 1934), 323–331.

[3] The discussion regarding the asymptotic behavior of $J_{m,n}(x)$ for large x will follow in a separate communication to this Journal.

The convergence of these two types of infinite series are discussed in §§2–3. An operational relation between a Humbert function and a Kummer function ${}_1F_1$ is deduced in §4. In the subsequent article an infinite series involving the product of a Humbert function and a Weber's parabolic cylinder function $D_n(x)$ is summed up in terms of Kelvin's function bei(x). Finally in §6 the summation of an infinite series involving a Humbert function and a Neumann's polynomial $0_n(t)$ are effected by means of parabolic cylinder functions.

II. Series of the Type (A)

From the relation (1), it is evident that for large n,

$$J_{m,n}(x) = \frac{(x/3)^n}{\Gamma(n+1)}[1 + O(n^{-1})]. \tag{4}$$

This by virtue of Stirling's formula

$$\Gamma(n) = n^{n-\frac{1}{2}}e^{-n}(2\pi)^{\frac{1}{2}}e^{\theta/12n} \tag{5}$$

gives that

$$J_{m,n}(x) = O\left(\frac{(x/3)^n}{n^{n+\frac{1}{2}}e^{-n}}\right) \tag{6}$$

for large n.

The series (A) is therefore convergent throughout the x-domain in which

$$\overline{\lim_{n\to\infty}} \sqrt[n]{\left|\left\{A_{m,n}\frac{(x/3)^n}{e^{-n}n^{n+\frac{1}{2}}}\right\}\right|} < 1.$$

In case the series (A) occurs in the form

$$\sum_{n=1}^{\infty} A_{m,n}\left(\frac{x}{3}\right)^{-n} J_{m,n}(x) \tag{A$_1$}$$

the necessary and sufficient condition for the convergence of (A_1) is that

$$\left|\frac{A_{m,n}}{e^{-n}n^{n-\frac{1}{2}}}\right| \tag{A$_1'$}$$

should tend to zero as n tends to infinity.

Example 1. The relation (1) can be written as

$$J_{m,n}(x) = \frac{(x/3)^{m+n}}{\Gamma(m+1)}\sum_{r=0}^{\infty}\frac{(-x^3/27)^r}{r!(m+1, r)\Gamma(n+r+1)}$$

where

$$(l, r) = l(l+1)(l+2)\cdots(l+r-1) = \frac{\Gamma(l+r)}{\Gamma(l)}.$$

This gives that

$$\sum_{n=0}^{\infty} \frac{(\alpha, n)(\beta, n)}{n!} \left(\frac{x}{3}\right)^{-n} J_{m,n}(x)$$

$$= \frac{(x/3)^m}{\Gamma(m+1)} \sum_{n=0}^{\infty} \frac{(\alpha, n)(\beta, n)}{n!} \sum_{r=0}^{\infty} \frac{(+x^3/27)^r}{(r!)^2(m+1, r)(r+1, n)}$$

$$= \frac{(x/3)^m}{\Gamma(m+1)} \sum_{r=0}^{\infty} \frac{(-x^3/27)^r}{(r!)^2(m+1, r)} \sum_{n=0}^{\infty} \frac{(\alpha, n)(\beta, n)}{n!(r+1, n)}$$

the inversion of the order of summation being justified on account of the absolute convergence of the two series involved.

But we know that

$$\sum_{n=0}^{\infty} \frac{(\alpha, n)(\beta, n)}{n!(r+1, n)} = {}_2F_1(\alpha, \beta; r+1; 1)$$

$$= \frac{\Gamma(r+1)\Gamma(r+1-\alpha-\beta)}{\Gamma(r+1-\alpha)\Gamma(r+1-\beta)}$$

$$R(r+1-\alpha-\beta) > 0.$$

Hence

$$\sum_{n=0}^{\infty} \frac{(\alpha, n)(\beta, n)}{n!} \left(\frac{x}{3}\right)^{-n} J_{m,n}(x)$$

$$(7) \qquad = \frac{(x/3)^m \Gamma(1-\alpha-\beta)}{\Gamma(1+m)\Gamma(1-\alpha)\Gamma(1-\beta)} \sum_{r=0}^{\infty} \frac{(1-\alpha-\beta, r)(-x^3/27)^r}{r!(m+1, r)(1-\alpha, r)(1-\beta, r)}$$

$$= \frac{(x/3)^m \Gamma(1-\alpha-\beta)}{\Gamma(1+m)\Gamma(1-\alpha)\Gamma(1-\beta)} {}_1F_3\left[1-\alpha-\beta; 1+m, 1-\alpha, 1-\beta; \frac{x^3}{27}\right]$$

provided that $R(\alpha+\beta) < 1$.

The series (7) is convergent, since the condition (A_1') here reduces after a little algebra, to $n^{\alpha+\beta-1}$, which on account of the condition $R(\alpha+\beta) < 1$, tends to zero as n tends to infinity.

The generalized hypergeometric series on the right of (7) reduces to Humbert functions when $m = -\alpha - \beta$ and when $\beta = 0$. Hence we obtain as special cases of (7), the following relations:

$$\sum_{n=0}^{\infty} \frac{(\alpha, n)(\beta, n)}{n!} \left(\frac{x}{3}\right)^{-n} J_{-\alpha-\beta,n}(x) = J_{-\alpha,-\beta}(x) \qquad R(\alpha+\beta) < 1$$

and

$$\sum_{n=0}^{\infty} \frac{(\alpha, n)}{n!} \left(\frac{x}{3}\right)^{-n} J_{m,n}(x) = \left(\frac{x}{3}\right)^{\alpha} \Gamma(1-\alpha) J_{m,-\alpha}(x) \qquad R(\alpha) < 1.$$

EXAMPLE 2. Another example of the series of the type (A) is the infinite series

$$(8) \qquad \sum_{n=0}^{\infty} \frac{J_{m,n}(-3x^2/4)}{n!} = (-x^2/4)^m \sum_{r=0}^{\infty} \frac{(x/2)^{5r}}{r!\Gamma(m+r+1)} J_r(x),$$

which can be easily established by the help of the definition (1), remembering also that

$$J_r(x) = \frac{(x/2)^r}{r!} {}_0F_1(r+1; -\tfrac{1}{4}x^2).$$

III. SERIES OF THE TYPE (B)

When $m = ln + k$, the series (1) reduces to

$$J_{ln+k,n}(x) = \frac{(x/3)^{ln+n+k}}{\Gamma(ln+n+k)\Gamma(k+1)} {}_0F_2\left(ln+k+1, n+1; -\frac{x^3}{27}\right).$$

This with the help of the estimate (5) gives that, for large n,

$$J_{ln+k,n}(x) = O\left(\frac{(x/3)^{ln+n}}{e^{-ln-n}\, l^{ln}\, n^{n+ln+k+1}}\right).$$

The series of the type (B) therefore converges throughout the x-domain in which

$$\overline{\lim_{n\to\infty}} \sqrt[n]{\left|\left\{A_{ln+k,n} \frac{(x/3)^{ln+n}}{e^{-ln-n}\, l^{ln}\, n^{n+ln+k+1}}\right\}\right|} < 1.$$

In case the series (B) is of the form

$$(B_1) \qquad \sum_{n=1}^{\infty} A_{ln+k,n} \left(\frac{x}{3}\right)^{-ln-n} J_{ln+k,n}(x)$$

the necessary and sufficient condition for the convergence will be that

$$(B_1') \qquad \left|\frac{A_{ln+k,n}}{e^{-ln-n}\, l^{ln}\, n^{n+ln+k}}\right|$$

should tend to zero as n tends to infinity.

EXAMPLE 1. Humbert[4] has investigated four series of the type. Of these we mention here

$$\sum_{n=0}^{\infty} \frac{(kx)^n}{3^n n!} J_{n,n}(x) = J_{o,o}(x\sqrt[3]{1+k})$$

from which, by putting $k = -1$, he has obtained the series

$$\sum_{n=0}^{\infty} \frac{(-x/3)^n}{n!} J_{n,n}(x) = 1.$$

[4] P. Humbert: Second paper quoted above, pp. 326–327. We can easily establish the convergence of the series investigated by Humbert by the help of the estimate of $J_{m,n}(x)$ given in this paper. There seems to be an omission of sign by Humbert in the series quoted here.

I propose to give now another generalization of this particular series by proving that

$$\sum_{n=0}^{\infty} \frac{(-x/3)^n}{n!} J_{n+\frac{1}{2}\nu, n+\frac{1}{2}\nu}(x) = \frac{(-)^{\frac{1}{2}\nu}(x/3)^\nu}{[\Gamma(\frac{1}{2}\nu + 1)]^2} \tag{9}$$

which for $\nu = 0$ reduces to Humbert's series just given above.

To establish (9), consider the known series[5]

$$\sum_{n=0}^{\infty} \frac{(\frac{1}{2}z)^{n+\frac{1}{2}\nu}}{n!} J_{n+\frac{1}{2}\nu}(z) = \frac{(z/2)^\nu}{\Gamma(\frac{1}{2}\nu + 1)}.$$

In this put $z = -2/\sqrt{p}$ where

$$\frac{1}{p\alpha} \doteqdot \frac{x^\alpha}{\Gamma(\alpha + 1)} \qquad R(\alpha + 1) > 0 \tag{10}$$

we then get

$$\sum_{n=0}^{\infty} \frac{(-)^n}{n!p^{\frac{1}{2}n+\frac{1}{2}\nu}} J_{n+\frac{1}{2}\nu}\left(-\frac{2}{\sqrt{p}}\right) = \frac{(-)^{\frac{1}{2}\nu}}{p^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu + 1)}.$$

Interpreting the left hand side by (2) and the right hand side by (10), then Lerch's theorem gives that

$$\sum_{n=0}^{\infty} \frac{(-)^n}{n!} x^{\frac{1}{2}n+\frac{1}{2}\nu} J_{n+\frac{1}{2}\nu, n+\frac{1}{2}\nu}(3\sqrt[3]{x}) = \frac{(-x)^{\frac{1}{2}\nu}}{[\Gamma(\frac{1}{2}\nu + 1)]^2}$$

which can be thrown in the form (9).

Example 2. An interesting series of the type (B) is given by

$$\sum_{r=0}^{\infty} \frac{(-\frac{1}{2}\sin 2\theta)^{\nu+2r}}{r!\Gamma(\nu + r + 1)} x^{\frac{1}{2}(\nu+2r)} J_{\nu+2r,\nu+2r}(3\sqrt[3]{x}) = J_{\frac{1}{2}\nu,\nu}(3\sqrt[3]{x\cos^2\theta}) J_{\frac{1}{2}\nu,\nu}(3\sqrt[3]{x\sin^2\theta}).$$

This series is important in as much as it gives the expansion of the product of two Humbert functions of different arguments in an infinite series involving Humbert functions.

Now if we put $m = \frac{1}{2}\nu$ and $n = \nu$, the image (2) reduces to

$$J_{\frac{1}{2}\nu,\nu}(3\sqrt[3]{x}) \doteqdot J_\nu(-2/\sqrt{p}).$$

If we use this and follow the method of example 1, this series can be easily established by interpreting both sides of the known series[6]

$$\sum_{r=0}^{\infty} \frac{(\frac{1}{2}z\sin 2\theta)^{\nu+2r}}{r!\Gamma(\nu + r + 1)} J_{\nu+2r}(z) = J_\nu(z\cos\theta)J_\nu(z\sin\theta).$$

A special case of the series investigated above is given by putting $\theta = \pi/4$, when we get that

$$\sum_{r=0}^{\infty} \frac{(-\frac{1}{2})^{\nu+2r}}{r!\Gamma(\nu + r + 1)} x^{\frac{1}{2}(\nu+2r)} J_{\nu+2r,\nu+2r}(3\sqrt[3]{x}) = [J_{\frac{1}{2}\nu,\nu}(3\sqrt[3]{x/2})]^2.$$

---

[5] Watson: Bessel Function, p. 525.

[6] W. N. Bailey: On the product of two Legendre Polynomials with different arguments, Proc. Lond. Math. Soc., (2), 41 (1936), 215–220.

IV. An Operational Image for Humbert Functions

A theorem[7] of operational calculus is that if

$$f(x) \doteqdot \phi(p)$$

then

$$f(x^2) \doteqdot \frac{p}{\sqrt{\pi}} \int_0^\infty e^{-\frac{1}{4}p^2s^2} \phi\left(\frac{1}{s^2}\right) ds. \tag{11}$$

Taking

$$f(x) = x^{\frac{2m-n}{3}} J_{m,n}(3\sqrt[3]{x})$$

and

$$\phi(p) = \frac{1}{p^{\frac{2m-n}{2}}} J_n\left(-2\sqrt{\frac{1}{p}}\right)$$

as given by (2), we get that

$$x^{\frac{4m-2n}{3}} J_{m,n}(3x^{2/3}) \doteqdot \frac{p}{\sqrt{\pi}} \int_0^\infty s^{2m-n} e^{-\frac{1}{4}p^2s^2} J_n(-2s)\, ds.$$

Since[8]

$$\int_0^\infty t^\lambda e^{-at^2} J_n(bt)\, dt = \frac{b^n \Gamma(\frac{1}{2}n + \frac{1}{2}\lambda + \frac{1}{2})}{2^{n+1} a^{\frac{1}{2}n+\frac{1}{2}\lambda+\frac{1}{2}} \Gamma(n+1)}$$

$$\times\, {}_1F_1\left(\tfrac{1}{2}n + \tfrac{1}{2}\lambda + \tfrac{1}{2}; n+1; -\frac{b^2}{4a^2}\right), \quad R(n+\lambda+1) > 0.$$

It follows that

$$x^{\frac{4m-2n}{3}} J_{m,m}(3x^{2/3}) \doteqdot \frac{(-)^n (4/p^2)^m \Gamma(m+\frac{1}{2})}{\sqrt{\pi}\Gamma(n+1)}\, {}_1F_1\left(m+\tfrac{1}{2}; n+1; -\frac{4}{p^2}\right) \tag{12}$$

$$R(m + \tfrac{1}{2}) > 0.$$

Similarly if we use the relation (3) in (11), we obtain that

$$x^{\frac{4n-2m}{3}} J_{m,n}(3x^{2/3}) \doteqdot \frac{(-)^m (4/p^2)^n \Gamma(n+\frac{1}{2})}{\sqrt{\pi}\Gamma(m+1)} \times {}_1F_1\left(n+\tfrac{1}{2}; m+1; -\frac{4}{p^2}\right). \tag{13}$$

The result (13) is also a consequence of (12), since $J_{m,n}(x)$ is symmetrical in m and n.

---

[7] P. Humbert: Le Calcul Symbolique (Paris, 1934), p. 28.

[8] Watson: Bessel Functions, p. 393.

3.2. We shall now show that there exist operational relations between Humbert functions and the various types of confluent hypergeometric functions.

Thus if we use Kummer's first transformation formula, viz.,

$$ {}_1F_1(\alpha;\rho;z) = e^z {}_1F_1(\rho-\alpha;\rho;-z) $$

(12) becomes

$$ \begin{aligned} x^{\frac{4m-2n}{3}} J_{m,n}(3x^{2/3}) &\doteqdot \frac{(-)^n(4/p^2)^m e^{-4/p^2}\Gamma(m+\frac{1}{2})}{\sqrt{\pi}\Gamma(n+1)} \\ &\qquad \times {}_1F_1\left(n-m+\tfrac{1}{2}; n+1; \frac{4}{p^2}\right). \end{aligned} \tag{14} $$

Since Whittaker's function $M_{k,m}(x)$ is defined by the relation

$$ M_{k,m}(x) = x^{m+\frac{1}{2}} e^{-\frac{1}{2}x} {}_1F_1(\tfrac{1}{2}+m-k; 2m+1; x) $$

(14) gives as a particular case, that

$$ x^{4/3r} J_{s+r,2s}(3x^{2/3}) \doteqdot \frac{(-)^s(4/p^2)^{r-\frac{1}{2}}\Gamma(s+r+\frac{1}{2})}{\sqrt{\pi}\Gamma(2s+1)} M_{r,s}(4/p^2) \quad R(s+r+\tfrac{1}{2}) > 0. $$

Taking appropriate values of m and n, we can similarly show that Humbert functions are operationally related to Laguerre polynomials $L_n^m(x)$, to Weber's parabolic cylinder functions $D_n(x)$, Bateman's functions $k_{2n}(x)$, and Bessel functions of the second kind $I_n(x)$.

V. An Infinite Series Involving the Product of a Humbert Function and a Parabolic Cylinder Function

If we put $b = \dfrac{2\sqrt{2}}{p}$ and $t = y\sqrt{2}$ in the known series[9]

$$ e^{-\frac{1}{2}t^2}\sin bt = e^{-\frac{1}{2}b^2}\sum_{n=0}^{\infty}\frac{(-)^n b^{2n+1}}{(2n+1)!} D_{2n+1}(t), $$

we get

$$ e^{-\frac{1}{2}y^2}\sin\frac{4y}{p} = \sum_{n=0}^{\infty}\frac{(-)^n 2^{3n+\frac{3}{2}} e^{-4/p^2}}{p^{2n+1}(2n+1)!} D_{2n+1}(y\sqrt{2}). \tag{15} $$

Consider p as a symbolic operator given by the relation (10). Putting $m = n + \frac{1}{2}$ in (12), we get that

$$ x^{\frac{2n+2}{3}} J_{n+\frac{1}{2},n}(3x^{2/3}) \doteqdot \frac{(-)^n}{\sqrt{\pi}} e^{-4/p^2}(4/p^2)^{n+\frac{1}{2}}. \tag{16} $$

If we now use (16) and the known operational image[10]

$$ \operatorname{bei}(2\sqrt{x}) \doteqdot \sin\frac{1}{p} $$

---

[9] R. S. Varma: On functions associated with the parabolic cylinder in harmonic analysis, Proc. Benares Math. Soc., 10 (1928), 15.

[10] Balth van der Pol: On the operational solution of linear differential equation and an investigation of the properties of these solutions, Phil. Mag. 8 (1929), 861–898.

and obtain the originals of both sides of (15), then Lerch's theorem gives that

$$(17)\qquad \sum_{n=0}^{\infty} \frac{2^n}{(2n+1)!} x^{2n/3} J_{n+\frac{1}{2},n}(3x^{2/3}) D_{2n+1}(y\sqrt{2}) = \frac{1}{\sqrt{(2\pi)}} x^{-2/3} e^{-\frac{1}{4}y^2} \operatorname{bei}(4\sqrt{xy}).$$

Since for large integral values[11] of n,

$$D_n(x) = \sqrt{2}(\sqrt{n})^n e^{-\frac{1}{2}n} \left[\cos(xn^{\frac{1}{2}} - \tfrac{1}{2}n\pi) + \frac{w_n(x)}{\sqrt{n}}\right]$$

where $w^n(x)$ satisfies both the inequalities

$$|w_n(x)| < \frac{3 \cdot 35 \ldots}{|x|\sqrt{\pi}} e^{\frac{1}{4}x^2}, \qquad |w_n(o)| < \tfrac{1}{8}n^{-\frac{1}{2}}$$

it is easy to see that the series (17) certainly converges, for all finite values of y within and on the circle $|x| = 1$.

VI. An Infinite Series Involving the Product of a Humbert Function and a Neumann's Polynomial

We consider the expansion

$$O_0(t)J_0(z) + 2\sum_{n=1}^{\infty} O_n(t)J_n(z) = \frac{1}{t-z} \qquad |z| < |t|.$$

Putting $z = -1/\sqrt{2p}$ and finding, in the manner of §5, the original of either side by the help of (10) and the known result[12]

$$(2x)^{\frac{1}{2}(m-1)} e^{\frac{1}{2}x} D_{-m}\{\sqrt{(2x)}\} \doteqdot \left(\frac{\pi}{2}\right)^{\frac{1}{2}} \frac{\sqrt{p}}{(\sqrt{p}+1)^m} \qquad R(m) > -1$$

we get that

$$(18)\qquad O_0(t)J_{0,0}(\tfrac{3}{2}\sqrt[3]{x}) + 2\sum_{n=1}^{\infty} O_n(t)J_{\frac{1}{2}n,n}(\tfrac{3}{2}\sqrt[3]{x}) = \left(\frac{\pi}{2}\right)^{-\frac{1}{2}} t^{-1} e^{x/4t^2} D_{-1}(\sqrt{x}/t).$$

If we use the relation[13]

$$O_n(t) = \frac{2^{n-1}n!}{t^{n+1}}\{1 + \phi_n\}$$

where $\phi_n \to 0$ as $n \to \infty$, we find that the region of convergence is confined to the domain given by

$$\overline{\lim_{n\to\infty}} \sqrt[n]{\left|\left\{\left(\frac{\sqrt{x}}{t}\right)^n \frac{e^{\frac{1}{2}n}}{n^{\frac{1}{2}n+\frac{1}{2}}}\right\}\right|} < 1.$$

Lucknow, India.

---

[11] Whittaker and Watson: Modern Analysis (Fourth Edition), p. 354.

[12] R. S. Varma: Summation of some infinite series of Weber's parabolic cylinder functions, Jour. Lond. Math. Soc., 12 (1937), 25-27.

[13] Whittaker and Watson: Modern Analysis (Fourth Edition), p. 375.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

ON TWO PROBLEMS OF SAMPLING

BY BROCKWAY MCMILLAN

(Received January 29, 1940)

I

We consider two problems which can be stated in the following terms: we have an arbitrarily large number of urns, each containing a mixture of white and black balls. If p_k is the density of white balls in the k^{th} urn (probability of drawing a white ball from the k^{th} urn), we suppose that the *a priori* probability that $p_k < x$ is given by $f_k(x)$, where

$$\begin{aligned} &f_k(x) = 0, \qquad x \leqq 0; \\ &f_k(x) = 1, \qquad x > 1; \\ &f_k(x) \leqq f_k(x+h), \qquad h \geqq 0; \\ &\lim_{h\to 0+} . f_k(x-h) = f_k(x). \end{aligned} \qquad k = 1, 2, \cdots \tag{1}$$

We suppose that these probabilities for separate urns are independent. We seek by taking a sample drawing of one ball from each of the first N urns to investigate the average density

$$\frac{p_1 + p_2 + \cdots p_N}{N}$$

of white balls in these urns. Our two different problems arise from two different methods of treating the sample.

For the asymptotic results which we wish to state, it will be necessary to think of the number of urns as being infinite, and to think of the sample as being taken from the whole array. The results of a drawing will be specified by an infinite sequence of indices, $(z_k, k = 1, 2, \cdots)$, defined by

$$z_k = 1 \text{ if white is drawn from the } k^{\text{th}} \text{ urn,}$$

$$z_k = 0 \text{ otherwise, } k = 1, 2, \cdots .$$

First problem: We suppose that the sample $(z_k, k = 1, 2, \cdots, N)$ is known exactly, and ask the *a posteriori* probability $Q_N((z_k), x)$ that

$$\frac{p_1 + p_2 + \cdots p_N}{N} < x. \tag{2}$$

Second problem: We suppose only that the density m/N of white balls in the sample is known, $m = z_1 + z_2 + \cdots z_N$, and ask the *a posteriori* probability $P_N(m/N, x)$ of (2).

We investigate the asymptotic behavior of $Q_N((z_k), x)$ and $P_N(m/N, x)$ for large N. When the functions $f_k(x)$ are all the same function, the analytical formulations of these two problems become identical. This simpler case has been treated by Bochner[1] and v. Mises[2]. Their results carry over to the more general problems when the moments of the $f_k(x)$ are sufficiently restricted. We have, in fact, under conditions to be stated below, that both $Q_N((z_k), x)$ and $P_N(m/N, x)$ tend with increasing N to a function $P(x)$ with the properties

$$(3) \qquad \begin{aligned} P(x) &= 0, \qquad x < p_0, \\ P(x) &= 1, \qquad x > p_0, \end{aligned}$$

where the "probable density" p_0 is defined in terms of the sample and the moments of the $f_k(x)$.

We suppose first that the moments

$$(4) \qquad a_k = \int_0^1 x\, df_k(x), \qquad k = 1, 2, \cdots,$$

are neither zero nor unity:

$$(5) \qquad a_k(1 - a_k) \neq 0, \qquad k = 1, 2, \cdots.$$

We next state a series of definitions, each for $k = 1, 2, \cdots$.

$$(6) \qquad \begin{aligned} g_k(x) &= \frac{1}{a_k} \int_{-\infty}^{x} p\, df_k(p) \\ h_k(x) &= \frac{1}{1 - a_k} \int_{-\infty}^{x} (1 - p)\, df_k(p) \end{aligned}$$

$$(7) \qquad \begin{aligned} b_{n,k} &= \int_0^1 x^n\, dg_k(x) \\ c_{n,k} &= \int_0^1 x^n\, dh_k(x) \end{aligned} \qquad n = 0, 1, 2, \cdots$$

$$(8) \qquad b_k = b_{1,k}, \qquad c_k = c_{1,k}$$

$$(9) \qquad d_k = b_{2,k} - (b_k)^2, \qquad e_k = c_{2,k} - (c_k)^2.$$

We note that the functions $g_k(x)$, $h_k(x)$ are distribution functions satisfying conditions of the form of (1). From this it follows that their dispersions (9) satisfy

$$(10) \qquad 0 \leqq d_k \leqq 1, \quad 0 \leqq e_k \leqq 1, \qquad k = 1, 2, \cdots.$$

We can now make more complete statements of the results indicated above. We recall that (5) is a restriction already imposed upon the $f_k(x)$.

---

[1] S. Bochner, "A Converse of Poisson's Theorem." Ann. Math. *37*, 1936, pp. 816–822.

[2] R. v. Mises, "A Modification of Bayes' Problem." Ann. Math. Stat. *9*, 1938, pp. 256–259.

(A) The sample (z_k) defines a sequence of distribution functions

$$F_k(x) = z_k g_k(x) + (1 - z_k)h_k(x), \qquad k = 1, 2, \cdots, \tag{11}$$

whose means r_k and dispersions s_k are given by

$$\begin{aligned} r_k &= z_k b_k + (1 - z_k)c_k \\ s_k &= z_k d_k + (1 - z_k)e_k . \end{aligned} \qquad k = 1, 2, \cdots \tag{12}$$

If the series $s_1 + s_2 + \cdots$ diverges, and if the sequence (r_k) has an average,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} r_k = r,$$

then $Q_N((z_k), x)$ tends as $N \to \infty$ to $P(x)$ of (3), with $p_0 = r$.

(B) Here we separate the restrictions on the sample (z_k) from those on the functions $f_k(x)$. We assume that the sample has a density:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} z_k = z. \tag{13}$$

Defining

$$l_k = \min. (d_k , e_k), \qquad k = 1, 2, \cdots, \tag{14}$$

we assume that the series $l_1 + l_2 + \cdots$ diverges. We further suppose that the sequences (b_k) and (c_k) possess averages in a certain strong sense: i.e., that there exist constants b and c such that[3]

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (b_k - b)^2 &= 0, \\ \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (c_k - c)^2 &= 0. \end{aligned} \tag{15}$$

Under these hypotheses, both $Q_N((z_k), x)$ and $P_N(m/N, x)$ tend to $P(x)$ of (3) with

$$p_0 = zb + (1 - z)c.$$

The strict asymptotic form of these results, as above stated, may be proved by an application of the Laplace-Liapounoff limit theorem, in the way that v. Mises[2] treats the simpler problem. We shall use the more powerful methods of Bochner[1], since they enable us to give estimates of the errors involved in replacing $Q_N((z_k), x)$ and $P_N(m/N, x)$ by $P(x)$. We shall, in fact, state our results without reference to limits as $N \to \infty$, in forms which will be valid for finite samples.

[3] Because the sequences in question are bounded, these are equivalent to conditions of the form $\lim N^{-1} \sum | b_k - b | = 0$.

II

This section reproduces essentially the argument of Bochner.[1] We assume (5). We consider a fixed sequence of zeros and ones, the sample (z_k, $k = 1, 2, \cdots, N$). Nearly all quantities with which we deal will depend upon the sample considered, but for simplicity of notation we shall not in this section attempt to indicate explicitly the fact of this dependence.

Let us assume for the moment that the sample is

$$z_k = 1, \qquad k = 1, 2, \cdots, m; \tag{16}$$
$$z_k = 0, \qquad k = m + 1, \cdots, N.$$

By Bayes' theorem, $Q_N((z_k), x) = Q_N(x)$ is given by

$$Q_N(x) = \frac{J_N(x)}{J_N(1)}$$

where $J_N(x)$ is the compound probability that simultaneously (2) shall hold and the drawing yield (16):

$$J_N(x) = \int_{p_1+p_2+\cdots+p_N<Nx} \cdots \int p_1 p_2 \cdots p_m(1 - p_{m+1}) \cdots (1 - p_N)\, df_1(p_1) \cdots df_N(p_N).$$

$J_N(1)$ can be calculated immediately. Recalling (4), we have

$$J_N(1) = a_1 a_2 \cdots a_m(1 - a_{m+1}) \cdots (1 - a_N). \tag{17}$$

In terms of the $F_k(x)$ defined by (11), $J_N(x)$ may be written

$$J_N(x) = a_1 \cdots a_m(1 - a_{m+1}) \cdots (1 - a_N) \int_{p_1+\cdots+p_N<Nx} \cdots \int dF_1(p_1) \cdots dF_N(p_N), \tag{18}$$

from which, by (17), we have

$$Q_N(x) = \int_{p_1+\cdots+p_N<Nx} \cdots \int dF_1(p_1) \cdots dF_N(p_N). \tag{19}$$

This last clearly holds independently of the temporary assumption (16).

It is known (Bochner[1]) that (19) may be expressed in the form

$$Q_N(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} E_N(w) \frac{\sin Nxw}{w}\, dw \tag{20}$$

where

$$E_N(w) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\,(iw(p_1 + p_2 + \cdots p_N))\, dF_1(p_1) \cdots dF_N(p_N).$$

This becomes

$$E_N(w) = \prod_{k=1}^{N} G_k(w)$$

where

$$G_k(w) = \int_0^1 e^{iwp}\, dF_k(p).$$

Consider the functions $\exp(-iwr_k)G_k(w)$, where r_k is defined by (12). Because

$$e^{-iwr_k} G_k(w) = \int_0^1 e^{iw(p-r_k)}\, dF_k(p),$$

we have the expansion

$$e^{-iwr_k} G_k(w) = \sum_{n=0}^{\infty} \frac{(iw)^n}{n!} G_{n,k}$$

valid for all w, with

$$G_{n,k} = \int_0^1 (x - r_k)^n\, dF_k(x), \qquad n = 0, 1, 2, \cdots, k = 1, 2, \cdots.$$

In particular, referring to (12), $G_{0,k} = 1$, $G_{1,k} = 0$, $G_{2,k} = s_k$. Furthermore, since $0 \leqq r_k \leqq 1$, for $n = 1, 2, \cdots$ we have

$$0 \leqq | G_{2+n,k} | \leqq \int_0^1 | x - r_k |^{2+n} dF_k(x) \leqq \int_0^1 | x - r_k |^2 dF_k(x)$$
$$= G_{2,k} = s_k \leqq 1.$$

Therefore

$$e^{-iwr_k}G_k(w) = 1 - s_k \frac{w^2}{2} + R_k(w) \tag{21}$$

where for $| w | \leqq 1$

$$| R_k(w) | \leqq s_k | w |^3 e. \tag{22}$$

By (10) we have then, uniformly in k

$$| R_k(w) | \leqq 3 | w |^3. \tag{23}$$

Defining

$$D(N) = s_1 + s_2 + \cdots + s_N,$$

we have from (21) and (22), for some w_0, $0 < w_0 \leqq 1$, and all w with $| w | \leqq w_0$, that

$$\log [e^{-iw(r_1+r_2+\cdots r_N)} E_N(w)] = -D(N) \frac{w^2}{2} + 0(w^3 D(N)).$$

From this, (23), and the identity

$$| U - V | = | U^{1/N} - V^{1/N} | \cdot | U^{(N-1)/N} + U^{(N-2)/N} V^{1/N} + \cdots V^{(N-1)/N} |$$

follows the existence of constants $A > 0$, $B > 0$, such that

$$| E_N(w)e^{-iw(r_1+r_2+\cdots r_N)} - e^{-\frac{1}{2}D(N)w^2} | \leqq AN \,| w |^3 e^{-BD(N)w^2} \tag{24}$$

holds for all w, $| w | \leqq w_0$, and all N.

Since we know only that $| E_N(w) | \leqq 1$, in place of (20) we consider the absolutely convergent integral

$$I_N(x) = \int_0^x Q_N(p)\, dp = \frac{2}{\pi} \int_{-\infty}^{\infty} E_N(w) \frac{(\sin \frac{1}{2}Nxw)^2}{Nw^2}\, dw.$$

The estimate (24) enables us to conclude that $I_N(x)$ differs from

$$\frac{2}{\pi} \int_{-\infty}^{\infty} [e^{iw(r_1+r_2+\cdots r_N)-\frac{1}{2}D(N)w^2}] \frac{(\sin \frac{1}{2}Nxw)^2}{Nw^2}\, dw \tag{25}$$

by

$$0\left[\int_{-w_0}^{w_0} | w | e^{-BD(N)w^2}\, dw\right] + 0\left[\int_{w_0}^{\infty} \frac{dw}{Nw^2}\right] = 0\left(\frac{1}{D(N)}\right) + 0\left(\frac{1}{N}\right).$$

Defining

$$r(N) = \frac{1}{N}(r_1 + r_2 + \cdots + r_N), \qquad H(N) = \frac{D(N)}{2N^2},$$

we rewrite (25) with $-w/N$ in place of w:

$$\frac{2}{\pi} \int_{-\infty}^{\infty} [e^{-iwr(N)-w^2H(N)}] \frac{(\sin \frac{1}{2}xw)^2}{w^2}\, dw. \tag{26}$$

This is equal to

$$\frac{1}{2\pi} \int_0^x dt \int_{-t}^{t} du \int_{-\infty}^{\infty} e^{iw(u-r(N))-w^2H(N)}\, dw$$

or

$$\frac{1}{2\sqrt{\pi H(N)}} \int_0^x dt \int_{-t}^{t} \exp\left(-\frac{(u-r(N))^2}{4H(N)}\right) du,$$

which becomes by integration by parts

$$\frac{x - r(N)}{2\sqrt{\pi H(N)}} \int_{-x-r(N)}^{x-r(N)} \exp\left(-\frac{u^2}{4H(N)}\right) du$$
$$+ \sqrt{\frac{H(N)}{\pi}} \left[\exp\left(-\frac{(x-r(N))^2}{4H(N)}\right) - \exp\left(-\frac{(x-r(N))^2}{4H(N)}\right)\right].$$

This differs from

$$\frac{x - r(N)}{2\sqrt{\pi H(N)}} \int_{-\infty}^{x-r(N)} \exp\left(-\frac{u^2}{4H(N)}\right) du$$

by a term which is majorized for $x \geqq 0$ by

$$0(\sqrt{H(N)}) + \frac{x - r(N)}{\sqrt{\pi}} \int_{\frac{x+r(N)}{2\sqrt{H(N)}}}^{\infty} e^{-u^2}\, du = 0(\sqrt{H(N)}) = 0\left(\frac{\sqrt{D(N)}}{N}\right).$$

In terms of the error function

$$\psi(x) = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{x} e^{-\frac{1}{4}u^2}\, du$$

we have altogether that for $x \geqq 0$

$$\left| I_N(x) - [x - r(N)] \cdot \psi\left[\frac{\sqrt{2}N(x - r(N))}{\sqrt{D(N)}}\right] \right| \leqq 0\left(\frac{\sqrt{D(N)}}{N}\right) + 0\left(\frac{1}{D(N)}\right). \tag{27}$$

We may use this to deduce result (A) stated above. By (10) we have that $D(N) \leqq N$. Defining

$$U(N) = |\, r - r(N) \,|,$$

we have from (27) that for $0 \leqq x \leqq \min\,(0, r - U(N))$

$$\int_0^x Q_N(p)\, dp \leqq 0\left(\frac{\sqrt{D(N)}}{N}\right) + 0\left(\frac{1}{D(N)}\right), \tag{29a}$$

while for $\min\,(1, r + U(N)) \leqq x \leqq 1$,

$$\left| \int_0^x Q_N(p)\, dp - (x - r) \right| \leqq U(N) + 0\left(\frac{\sqrt{D(N)}}{N}\right) + 0\left(\frac{1}{D(N)}\right). \tag{29b}$$

These and (10) imply that for $h > U(N)$

$$\begin{aligned} Q_N(r - h) &\leqq \frac{1}{h}\left[0\left(\frac{1}{\sqrt{N}}\right) + 0\left(\frac{1}{D(N)}\right)\right] \\ |\, Q_N(r + h) - 1 \,| &\leqq \frac{1}{h}\left[U(N) + 0\left(\frac{1}{\sqrt{N}}\right) + 0\left(\frac{1}{D(N)}\right)\right], \end{aligned} \tag{30}$$

since $Q_N(x)$ is non-decreasing in x. This establishes result (A) when $r = \lim r(N)$.

Using (24), with further restrictions on the $f_k(x)$, we may state an estimate analogous to (27) for $Q_N(x)$ itself. Essentially, all we require is that

$$\left| \int_{w_0}^{\infty} + \int_{-\infty}^{-w_0} E_N(w) \frac{\sin Nxw}{w}\, dw \right| \tag{31}$$

vanish as $N \to \infty$, for any x, $0 \leqq x \leqq 1$, and any w_0, $0 < w_0 \leqq 1$. If, for example, the distribution functions $f_k(x)$ have densities $f_k'(x)$ of variations bounded uniformly in k, (31) is dominated by $0(N^{-1})$. We suppose that for

any $w_0 > 0$ there is a function $V(N)$ which dominates (31) uniformly in $0 \leqq x \leqq 1$. We have by (24) then that $Q_N(x)$ differs from

$$\frac{1}{\pi}\int_{-\infty}^{\infty} [e^{iwNr(N)-\frac{1}{2}D(N)w^2}] \frac{\sin Nxw}{w}\, dw \tag{32}$$

by a term which is dominated by

$$0\left[N\int_{-w_0}^{w_0} w^2 e^{-BD(N)w^2}\,dw\right] + V(N) + 0\left[\int_{w_0}^{\infty} e^{-\frac{1}{2}D(N)w^2}\frac{dw}{w}\right]$$
$$= 0(ND(N)^{-\frac{1}{2}}) + V(N).$$

The integral (32) can be written

$$\frac{1}{2\pi}\int_{-x}^{x} du \int_{-\infty}^{\infty} e^{iw(u-r(N))-H(N)w^2}\,dw$$

or

$$\frac{1}{2\sqrt{\pi H(N)}}\int_{-x-r(N)}^{x-r(N)} \exp\left(-\frac{u^2}{4H(N)}\right) du.$$

If $r(N) \geqq \epsilon > 0$ for all N, this last differs from

$$\psi\left[\frac{\sqrt{2}N(x - r(N))}{\sqrt{D(N)}}\right]$$

by

$$0\left[\int_{-\infty}^{-\frac{\epsilon}{\sqrt{H(N)}}} e^{-\frac{1}{2}u^2}\,du\right] \leqq 0[\sqrt{H(N)}e^{-\frac{1}{2}\epsilon^2/H(N)}]$$

uniformly in $0 \leqq x \leqq 1$. Under the assumption, then, that $r(N) \geqq \epsilon > 0$ for large N, we have

$$\left|Q_N(x) - \psi\left[\frac{\sqrt{2}N(x - r(N))}{\sqrt{D(N)}}\right]\right| \leqq V(N) + 0(ND(N)^{-\frac{1}{2}}). \tag{33}$$

III

We now turn to our second problem. Again we assume (5). For convenience, we shall speak of a sequence (z_k, $k = 1, 2, \cdots, N$) such that $z_1 + z_2 + \cdots + z_N = m$ as an m-sequence. We seek the probability $P_N(m/N, x)$ of (2) when it is known that an m-sequence has been drawn. By Bayes' theorem

$$P_N(m/N, x) = \frac{K_{m,N}(x)}{K_{m,N}(1)}$$

where $K_{m,N}(x)$ is the probability of the drawing of an m-sequence simultaneously with the occurrence of (2). That is, using the notation of the last section,

$$K_{m,N}(x) = \sum J_N((z_k), x) = \sum J_N((z_k), 1)Q_N((z_k), x)$$

where the sum is taken over all the C_m^N possible m-sequences. We have, therefore,

$$P_N(m/N, x) = \frac{\sum J_N((z_k), 1)Q_N((z_k), x)}{\sum J_N((z_k), 1)},$$

which exhibits $P_N(m/N, x)$ as a weighted average over all distinct m-sequences of the functions $Q_N((z_k), x)$ studied in the preceeding section. The results of that section will carry over, then to $P_N(m/N, x)$ once certain uniformity requirements have been met.[4]

Recalling (14), we define

$$L(N) = \sum_{k=1}^{N} l_k \leqq \min_{z_1+z_2+\cdots+z_N=m} [D((z_k), N)]$$

and

$$T(N) = \max_{z_1+z_2+\cdots z_N=m} | p_0 - r((z_k), N) |.$$

We have from (30) that for $h \geqq T(N)$

$$(34)\qquad \begin{aligned} P_N(m/N, p_0 - h) &\leqq \frac{1}{h} [0(N^{-\frac{1}{2}}) + 0(L(N)^{-1})] \\ | P_N(m/N, p_0 + h) - 1 | &\leqq \frac{1}{h} [T(N) + 0(N^{-\frac{1}{2}}) + 0(L(N)^{-1})]. \end{aligned}$$

This, together with (30), will establish result (B). For any m-sequence and $p_0 = zb + (1 - z)c$ (see (13), (15))

$$\begin{aligned} | r((z_k), N) - p_0 | &= \left| \frac{1}{N} \sum_{k=1}^{N} [z_k(b_k - b) + (1 - z_k)(c_k - c)] \right. \\ &\qquad \left. + \frac{m}{N} b + \left(1 - \frac{m}{N}\right) c - p_0 \right| \\ &\leqq \frac{1}{N} \sum_{k=1}^{N} [| b_k - b | + | c_k - c |] + \left| \frac{m}{N} - z \right| (b + c) \\ &\leqq \left[\frac{1}{N} \sum_{k=1}^{N} (b_k - b)^2\right]^{\frac{1}{2}} + \left[\frac{1}{N} \sum_{k=1}^{N} (c_k - c)^2\right]^{\frac{1}{2}} + 2\left|\frac{m}{N} - z\right|, \end{aligned}$$

the last step by virtue of the Schwarz inequality. From this, we see that the hypotheses of (B) insure that $T(N) \rightarrow 0$; since they also insure that $L(N) \rightarrow \infty$, the result follows from (30) and (34).

An estimate anlogous to (33) for $P_N(m/N, x)$ is not possible without a restriction on the rate at which the limit in (13) is approached, and consequently has no meaning for finite samples. In this connection, it might be pointed out that the use of the error function $\psi(x)$ in stating, for example, inequalities (27) and (33) is largely a concession to convention. The right members of these inequalities would remain unchanged if $\psi(x)$ were replaced by a step function analogous to $P(x)$ of (3).

PRINCETON UNIVERSITY, PRINCETON, N. J.

[4] It is, in fact, for this reason that our first problem was introduced. All our attempts to apply directly to $P_N(m/N, x)$ analysis analogous to that in section II above bogged down in a morass of combinatory coefficients.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

TRANSFORMATIONS OF FINITE PERIOD. III

Newman's Theorem

BY P. A. SMITH

(Received February 22, 1940)

We shall give a new proof of a theorem of M. H. A. Newman [3] which asserts that periodic transformations of given period p operating in a locally euclidean space M can not be arbitrarily "small" (where by a small transformation is meant one which displaces points by a uniformly small amount). Our proof is quite different from Newman's and in a sense more direct; for only the space M which is being transformed comes into consideration whereas Newman's proof has for its setting the topological product of p copies of M. The methods which we shall use enable us to dispense with the locally euclidean restriction and to obtain a proof for far more general spaces. Moreover we establish a theorem which is slightly stronger than Newman's since it asserts, for periodic transformations operating in M, that there exists an impossible degree of smallness which is independent of the period. In this result, however, smallness must be understood in terms of orbits rather than displacements. To these considerations which in themselves perhaps justify a new proof of a known theorem, we add the following. The question whether or not there can exist *any* group—not merely finite, cyclic—operating effectively in a reasonably regular space and defining uniformly small orbits, is an outstanding problem in the theory of topological groups[1] and new methods of treating the special case under consideration will perhaps not be without interest.

A SPECIAL CASE

1. In order to make clear the nature of the theorem to be proved and to reveal the underlying ideas of the proof, which might otherwise remain hidden by the details, we shall first consider a simple example.

Let T be a transformation operating in a space M,—that is a one-one bicontinuous transformation of M into itself. The totality of images of a given point x of M under positive and negative powers of T constitute the *orbit* of x defined by T.

Let M_2 be an ordinary 2-sphere in euclidean 3-space. We shall call a *cap* of M_2 any closed circular region on M_2 which is smaller than a hemisphere. The special theorem to be proved is the following:

Every periodic transformation operating in M_2 defines at least one orbit which is not contained in any cap.

[1] For example, a negative answer to this question would establish the existence of one-parameter subgroups in locally euclidean topological groups.

Proof. Suppose on the contrary that there exists a transformation T of period q operating in M_2 such that every orbit defined by T is contained in a cap. It is easy to see that for each point x of M_2 there will be a uniquely determined smallest cap $u(x)$ containing the orbit of x defined by T and that the center $\omega(x)$ of $u(x)$ will be a continuous function of x.

We shall make use of the well known fact that every periodic transformation operating in a sphere is topologically equivalent to a periodic orthogonal transformation. From this it follows that there can be introduced into M_2 a triangulation which is preserved by T and whose 2-cells can be represented without repetition by

$$E^i, TE^i, \cdots, T^{q-1}E^i \qquad (i = 1, \cdots, \alpha).$$

We may assume that each E^i is positively oriented relative to a definitely chosen orientation of M. Let us assume for definiteness that T preserves orientation. Then the chain

$$\Delta = \sum_i (E^i + TE^i + \cdots + T^{q-1}E^i)$$

is, relative to every coefficient group, a fundamental 2-cycle on M_2 and as such, can not be ~ 0 on M_2. In particular Δ can not be $\sim 0 \bmod q$.

Let μ denote the single-valued continuous mapping $x \to \omega(x)$ defined over M_2; it carries Δ into a singular 2-cycle Δ^* on M. Moreover μ can be obtained by performing a deformation: we have merely to slide x to $\omega(x)$ along that geodesic arc which is shorter than a great semicircle, and at a suitable rate of speed depending on x. It follows that $\Delta \sim \Delta^*$ on M_2. Now from the way μ is defined, it is clear that $\mu(x) = \mu(Tx)$ and therefore that $\mu E^i = \mu TE^i$. Consequently

$$\Delta^* = \mu(\Sigma(E^i + \cdots + T^{q-1}E^i) = q\mu(\Sigma E^i) = 0 \bmod q$$

so that $\Delta \sim 0 \bmod q$, which is impossible. This completes the proof for the case in which T preserves orientation; minor modifications yield a proof for the orientation-reversing case.

2. Suppose that the transformation T of period q operating in M_2 admits at least one fixed point, say x_0. Then for points x near x_0, the functions $u(x)$ and $\omega(x)$ introduced above are uniquely defined and $\omega(x)$ is continuous. Let x vary continuously along a path joining x_0 to a point say x_1 different from x_0. Then $\omega(x)$ will remain defined and continuous so long as the orbit of x is contained in a cap (the cap depending of course on x). As a consequence of the theorem above, $\omega(x)$ is not defined for all x. Suppose x_1 is a point for which $\omega(x)$ is not defined. Then as x moves along the path x_0x_1, we must come to a last point $\bar{x}$ whose orbit is contained in a cap. From elementary continuity consideration, however, it is fairly clear that the orbit of $\bar{x}$ will be contained in a hemisphere H and that at least two points of the orbit of $\bar{x}$ lie on the great

circle boundary of H; if there are exactly two, they must be diametrically opposite each other.

Suppose again that T is a transformation of period q operating in M_2. It can be shown without great difficulty that if the geodesic distance $d(x, Tx)$ is uniformly smaller than $2\pi r/q$, r being the radius of M_2, then each orbit defined by T would be contained in a cap. Since this is impossible, we conclude that there exists at least one point x such that $d(x, Tx) \geqq 2\pi r/q$. In particular, if T admits fixed points, there exists at least one x such that $d(x, Tx) = 2\pi r/q$.

Although these results can be generalized to higher dimensions, a number of technical details enter the proof due to the fact that for spaces of higher dimensions, there is no known method of constructing invariant triangulations. It will be seen in what follows how this difficulty can be met. We shall not, however, pay further attention to transformations of spheres but merely state the results:

Let M_n be an n-sphere of radius r in euclidean $(n + 1)$-space. Every periodic transformation operating in M_n admits at least one orbit which is not contained in any cap (a cap being any closed region on M_n bounded by an $(n - 1)$-sphere which is smaller than a great $(n - 1)$-sphere). *Suppose T is a transformation of period q operating in M_n. If T admits fixed points, there exists at least one point x such that $d(x, Tx) = 2\pi r/q$ where d denotes geodesic distance. Moreover, there will exist at least one point y and integer $s(2 \leqq s \leqq q)$ such that s points of the orbit of y lie on some great $(s - 2)$-sphere of M* (a great 0-sphere being a pair of diametrically opposite points).[2]

More General Spaces

3. From now on the word space will mean a Hausdorff space in which every open set is the sum of a countable family of closed sets (Cf. [3], p. 134).

We shall consistently use the letter M to denote a space, T a transformation operating in M, N a set of points in M and $\mathfrak{A}$ a covering of M (i.e. a finite covering by open sets). We shall write "$T < \mathfrak{A}$ over N" if to each point x of N there can be associated a set of $A_x \in \mathfrak{A}$ such that $x + Tx \subset A_x$. We shall write "$T \ll \mathfrak{A}$ over N" if to each x in N there can be associated an $A_x \in \mathfrak{A}$ which contains the orbit of x.

Lemma 1. Suppose M is locally bicompact, N bounded.[3] For a given integer q and covering $\mathfrak{A}$ there exists a covering $\mathfrak{A}'$ such that every periodic T of period q which satisfies the relation $T < \mathfrak{A}'$ over N satisfies also the relation $T \ll \mathfrak{A}$ over N.

Lemma 2. Suppose M is locally bicompact and H, K are bounded open

---

[2] The methods suggested by the example we have considered can also be made to yield the theorem that the orbits of a periodic transformation operating in euclidean n-space cannot be uniformly bounded.

[3] A set will be called bounded if its closure is bicompact.

sets such that $\bar{K} \subset H$. There exists a covering $\mathfrak{A}$ such that if $T \ll \mathfrak{A}$ over N (T not necessarily periodic) then

$$\sum_n T^n \bar{K} \subset H.$$

The proofs of lemmas 1 and 2 are elementary.

4. Regularity. Cycles and homologies are to be understood in the sense of Čech (see [2]). Let M be a space, $\mathfrak{g}$ a coefficient group for cycles and homologies in M and n an integer $\geqq 0$. We shall say that M possesses property p_n over $\mathfrak{g}$ at a point d if there exists a neighborhood $D(d)$ and an n-cycle Ω mod $M - D$ with coefficients in $\mathfrak{g}$ such that (1) for no neighborhood $D'(d)$ contained in D is $\Omega \sim 0$ mod $M - D'$ (over $\mathfrak{g}$); (2) for every neighborhood $B(d) \subset D$, there exists a neighborhood $B'(d) \subset B$ such that over $\mathfrak{g}$, Ω is a basis for n-cycles mod $M - B$ relative to homologies mod $M - B'$ while (3) cycles mod $M - B$ of dimension exceeding n are ~ 0 mod $M - B'$. A neighborhood $D(d)$ and relative cycle Ω satisfying (1) (2) (3) will be said to constitute an n-dimensional *fundamental pair* (Ω, D) for d over $\mathfrak{g}$. Clearly if (Ω, D) is an n-dimensional fundamental pair for d over $\mathfrak{g}$, then so is the pair obtained by replacing D by a smaller neighborhood of d.

We shall say that M possesses property q_n over $\mathfrak{g}$ at d if for every neighborhood $A(d)$ there exists a neighborhood $A'(d) \subset A$ such that if C is a non-empty open subset of A', every n-cycle mod $M - A$ in $M - C$ is ~ 0 mod $M - A'$ (all coefficients being in $\mathfrak{g}$).

It will be convenient to call M *n-regular* over $\mathfrak{g}$, if it possesses properties p_n and q_n at each of its points.[4]

We can now state Newman's theorem for spaces of a very general type.

THEOREM I. *Let M be a connected locally bicompact finite dimensional space, N a bounded open set in M and q an integer $\geqq 2$. Let p be a prime factor of q and $\mathfrak{g}_p$ the additive group of integers reduced modulo p. If M is n-regular over $\mathfrak{g}_p$, there exists a covering $\mathfrak{A}$ of M such that no periodic transformation T of period q operating in M can satisfy the relation $T < \mathfrak{A}$ over N.*

The covering $\mathfrak{A}$ necessarily depends on q as one can see from simple examples. But if one replaces the relation $T < \mathfrak{A}$ by the relation $T \ll \mathfrak{A}$ and sharpens the hypothesis on M by requiring that regularity possess a certain uniformity relative $\mathfrak{g}_2, \mathfrak{g}_3, \mathfrak{g}_5, \cdots$ then there can be chosen an $\mathfrak{A}$ independent of q. We shall make this more precise. Suppose that there exist a neighborhood D of d such that for every prime p there exist an n-dimensional fundamental pair (Ω^p, D) over $\mathfrak{g}_p$ for d. We shall say that the fundamental pairs (Ω^p, D), $p = 2, 3, 5, \cdots$, constitute a *fundamental family* $F(D, d, n)$.

Let $F(D, d, n)$ be a fundamental family of fundamental pairs (Ω^p, D). For each p, conditions (1) (2) (3) in the definition of fundamental pair, are satisfied

[4] Properties p_n and q_n are slightly weaker than P_n, Q_n in our paper [4].

relative to $\mathfrak{g}_p$. We shall say that conditions (2) and (3) are satisfied *uniformly* by $F(D, d, n)$ if B' is independent of B,—that is, if for every $B(d)$ there is a $B'(d) \subset B$, B' independent of p, such that for every p, Ω^p is a basis over $\mathfrak{g}_p$ for n-cycles mod $M - B$ relative to homologies mod $M - B'$ while cycles mod $M - B$ over $\mathfrak{g}_p$ of dimensions exceeding n are 0 mod $M - B'$. Consider a neighborhood $D'(d) \subset D$. Then for each p, $\Omega^p \not\sim 0$ mod $M - D'$ over $\mathfrak{g}_p$ (condition (1)). Hence for each p there exists a covering $\mathfrak{U}$ of M such that

$$\text{(A)} \qquad \Omega^p(\mathfrak{U}) \not\sim 0 \text{ mod } M - D' \text{ over } \mathfrak{g}_p .$$

If for a given D' there exists a $\mathfrak{U}$ independent of p such that (A) is satisfied for each p, we shall say that $F(D, d, n)$ satisfies condition (1) uniformly. A fundamental family which satisfies (1) (2) and (3) uniformly will be called a *uniform family*. It is easy to see that if $F(D, d, n)$ is a uniform family of fundamental pairs (Ω^p, D), then the fundamental pairs obtained by replacing D by a smaller neighborhood D_1 of d constitute a uniform family $F(D_1, d, n)$.

If for every point d of M there exists a uniform family $F(D, d, n)$, $D = D(d)$, we shall say that property p_n is satisfied uniformly in M. Similarly, property q_n is satisfied uniformly if for each point d and neighborhood $A(d)$, there exists an $A'(d)$ such that if C is a non-null open subset of A and p an arbitrary prime, n-cycles mod $M - A$ in $M - C$ with coefficients in $\mathfrak{g}_p$ are ~ 0, mod $M - A$ (over $\mathfrak{g}_p$). We shall say that M is uniformly n-regular with respect to arbitrary moduli if properties p_n and q_n are satisfied uniformly in M. It can be shown without great difficulty that locally euclidean n-dimensional spaces are uniformly n-regular.

THEOREM II. *Let M be a locally bicompact finite dimensional space and N a bounded open set in M. If M is uniformly n-regular with respect to arbitrary moduli, there exists a covering $\mathfrak{A}$ of M such that no periodic transformation T operating in M can satisfy the relation $T \ll \mathfrak{A}$ over N.*

With regard to theorem I, it is sufficient to prove it for the case in which $q = p =$ a prime. For suppose that $q = hp$, p a prime, and that M is n-regular over $\mathfrak{g}_p$. Then if T is of period q, T^h is of period p and therefore, if $\mathfrak{A}$ is suitably chosen, T can not satisfy the relation $T^h < \mathfrak{A}$ over N. By lemma 1 there exists a covering $\mathfrak{A}'$ such that if $T < \mathfrak{A}'$ over N, then $T \ll \mathfrak{A}$ over N. Since the second of these relations would imply that $T^h < \mathfrak{A}$ over N, T can not satisfy the first. Even simpler considerations make it clear that in order to prove theorem II we need only show that an $\mathfrak{A}$ exists such that no transformation of *prime* period can satisfy the relation $T \ll \mathfrak{A}$ over N.

We remark further that were it not for the fact that in theorem I we assume regularity relative to a single group $\mathfrak{g}_p$, theorem I would be a consequence of theorem II and lemma 1. But it will be seen in the proof of theorem II that in proving the non-existence of small transformations of any *given* prime period p, only the hypothesis of regularity over $\mathfrak{g}_p$ is used. Thus the proof of theorem I is contained essentially in that of theorem II and will therefore not be given separately.

5. It will simplify matters considerably if we carry out the proof of theorem II under the assumption that M is bicompact and then indicate the modifications which will extend the proof to the locally bicompact case. Let us assume then that M is a bicompact space satisfying the hypotheses of theorem II, and that N is an open set in M. Let $m = \dim M$; evidently[5] $m \geqq n$.

Let T be a periodic transformation of prime period p operating in M. We shall denote by L^T the totality of points of M which are left fixed by T, and by $\{\mathfrak{U}^T\}$ the totality of *special* coverings ([3], page 136) of M relative to T. For our present purposes, the essential properties of a special covering $\mathfrak{U}^T$ are the following: (1) the component sets of $\mathfrak{U}^T$ are permuted among themselves by T so that T induces a simplicial transformation of the *complex* (nerve of) $\mathfrak{U}^T$ into itself; (2) the simplexes of $\mathfrak{U}^T$ which are invariant under T form a subcomplex $\mathfrak{U}_0^T$; (3) the simplexes of $\mathfrak{U}_0^T$, and these only, are in[6] L^T; (4) each simplex in $\mathfrak{U}^T - \mathfrak{U}_0^T$ is of dimension $\leqq m$. It is evident that a necessary and sufficient condition that a $\mathfrak{U}^T$-chain be in L^T (in the sense of footnote 5) is that it be in the subcomplex $\mathfrak{U}_0^T$. Moreover, every chain in $\mathfrak{U}^T - \mathfrak{U}_0^T$ is in $M - L^T$, but the converse is not necessarily true. The totality of coverings $\mathfrak{U}^T$ is a complete system ([3], p. 137) and therefore can be used for defining all homology relations in M.

A transformation T of prime period p operating in M induces a boundary-preserving operation (also denoted by T) on the $\mathfrak{U}^T$-chains in M. Let δ, σ denote the operators $1 - T$ and $1 + T + \cdots + T^{p-1}$. We shall use the symbols ρ, $\bar{\rho}$ to stand either for δ, σ or σ, δ respectively. When the coefficient group $\mathfrak{g}$ is $\mathfrak{g}_p$, all $\mathfrak{U}^T$-chains in $\mathfrak{U}_0^T$ are annihilated by ρ and $\bar{\rho}$. For if E is a $\mathfrak{U}_0^T$-simplex, then $\delta E = E - E = 0$ and $\sigma E = pE = 0 \pmod p$. From this remark follows

LEMMA 3. If $\mathfrak{g} = \mathfrak{g}_p$ where p is the period of T, assumed to be prime, then every $\mathfrak{U}^T$-chain of the form ρX (or $\bar{\rho}X$) is in $\mathfrak{U}^T - \mathfrak{U}_0^T$, hence in $M - L$.

Suppose now that B is an invariant set, that is, a set which coincides with its image under T. Suppose there exists a $\mathfrak{U}^T$-chain of the form ρX which is a cycle mod B. If ρX is, modulo B, the boundary of a $\mathfrak{U}^T$-chain of the same form, we shall write $\rho X \simeq 0 \bmod B$. In particular, if $\rho X = 0 \bmod B$, then $\rho X \simeq 0 \bmod B$.

[5] The regularity hypotheses on M imply that for each p the modulo p dimension (Alexandroff) of M is n. This however, so far as we know, does not imply that $\dim M = n$. If it were definitely assumed that $\dim M$ equals n (not merely that it is finite), the proof of Theorem II would be somewhat shorter since the existence of the relative cycles in propositions A), C), and F) could be established without the use of lemma 4. We prefer to consider the more general situation on account of possible connections with the problem mentioned in the introduction. For in this problem one is led to consider transformations in spaces about whose dimension nothing is as yet known.

[6] A $\mathfrak{U}$-simplex $(U_0U_1 \cdots U_h)$ is in the set A if $U_0U_1 \cdots U_hA \neq 0$. A $\mathfrak{U}$-chain is in A if each of its simplexes is in A. A cycle is in A if each of its coordinates is in A.

LEMMA 4. Let $\mathfrak{g} = \mathfrak{g}_p$ and suppose that ρX_h, $\bar{\rho}X_{h-1}$ are $\mathfrak{U}^T$-cycles mod B such that

$$\beta X_h = \bar{\rho} X_{h-1} \bmod B$$

(β being the boundary operator). If $\rho X_h \simeq 0 \bmod B$, then $\bar{\rho}X_{h-1} \simeq 0$ mod B.

Proof. Suppose in fact that there exists a relation of the form $\beta\rho X_{h+1} = \rho X_h \bmod B$. Write

$$(1) \qquad \beta X_{h+1} = X_h - Z.$$

On operating on both sides of (1) by ρ we find that $\rho Z = 0 \bmod B$, hence by [3], section 3.11, we may write $Z = \bar{\rho}W + W^L \bmod B$ with $W^L \subset L$. Therefore, if we operate on both sides of (1) by β we obtain

$$0 = P + Q \bmod B, \qquad P = \bar{\rho}(X_{h-1} - \beta W), \qquad Q = \beta W^L.$$

Since Q is in $\mathfrak{U}_0^T$ and since, by lemma 3, P is in $\mathfrak{U}^T - \mathfrak{U}_0^T$, the chains P and Q have no common simplex, hence $P = Q = 0 \bmod B$. In particular $\beta\bar{\rho}W = \bar{\rho}X_{h-1} \bmod B$ so that $\bar{\rho}X_{h-1} \simeq 0 \bmod B$.

6. We now prove[7] theorem II for the bicompact case by establishing a sequence of propositions concerning chains and cycles associated with transformations of prime period. Let M be a connected bicompact space, uniformly n-regular relative to arbitrary moduli. Let N be an open subset of M and let $m = \dim M$. Let it be understood in propositions A) B) C) D) that T is an arbitrary but fixed transformation of prime period p operating in M, and that the coefficient group is $\mathfrak{g}_p$.

A) Let d be a point of M. Suppose there exists for d an n-dimensional fundamental pair (Ω, G) such that G is invariant under T. Let $k = m - n + 1$ and suppose further that there exist invariant neighborhoods of d:

$$G \supset A_0 \supset A_1 \supset \cdots \supset A_k$$

such that (1) Ω is a basis for n-cycles mod $M - G$ relative to homologies mod $M - A_0$; (2) for $i = 0, \cdots, k$, cycles mod $M - A_{i-1}$ of dimension $> n$ are ~ 0 mod $M - A_i$. Then there exists an n-cycle Δ mod $M - A_k$ of the form

$$\{\sigma\Gamma(\mathfrak{U}^T) + \Gamma^L(\mathfrak{U}^T)\}, \qquad \Gamma^L(\mathfrak{U}^T) \subset L^T$$

such that $\Delta \sim \Omega \bmod M - A_k$.

Proof. Since G is invariant, $T\Omega$ is a cycle mod $M - G$ and hence $T\Omega \sim x\Omega$ mod $M - A_0$ where $x \in \mathfrak{g}_p$. Since A_0 is also invariant, this homology holds if $T\Omega$ is substituted for Ω. On repeated substitution, we obtain: $\Omega = T^p\Omega \sim x^p\Omega \bmod M - A_0$, hence $x^p = 1$, $x = 1$, so that

$$(1) \qquad \delta\Omega \sim 0 \bmod M - A_0.$$

[7] The proof given here is somewhat shorter and simpler than an earlier proof proposed by the author, thanks to valuable suggestions from Professor Norman Steenrod.

Assume for the moment that $m > n$. Let $\mathfrak{U}^T$ be a special covering and i an integer in the range $1, \cdots, k$ and h an integer in the range $n + 1, \cdots, m$. From the Čech homology theory, there exists a special refinement $\mathfrak{B}^T$ of $\mathfrak{U}^T$ such that the projection into $\mathfrak{U}^T$ of any h-dimensional $\mathfrak{B}^T$-cycle mod $M - A_{i-1}$ will be the $\mathfrak{U}^T$-coordinate of a Čech cycle mod $M - A_{i-1}$, hence will (by the definition of the A's) be ~ 0 mod $M - A_i$ (see [2]). It is obvious that $\mathfrak{B}^T$ can be chosen to be independent of h since the range of h is finite.—Now let $\mathfrak{U}^T$ be an arbitrarily chosen special covering. From what has been said, we may choose special coverings

$$\mathfrak{U}^T = \mathfrak{U}_k \supset \mathfrak{U}_{k-1} \supset \cdots \supset \mathfrak{U}_0$$

such that for $i = 1, \cdots, k$, the projection into $\mathfrak{B}_i$ of a $\mathfrak{U}_{i-1}$-cycle mod $M - A_{i-1}$ of dimension exceeding n but not m, is ~ 0 mod $M - A_i$. Let π_i be a projection $\mathfrak{U}_{i-1} \to \mathfrak{U}_i$ which is permutable with T,—it is easy to see that such projections exist (see [3], p. 138). Let $\rho_0, \rho_1, \cdots$ stand alternately for δ, σ beginning with $\rho_0 = \delta$.

Putting $\Omega = \Omega_n$, it follows from (1) that $\rho_0\Omega_n(\mathfrak{U}_0) \sim 0$ mod $M - A_0$. Hence $\pi_1\rho_0\Omega_n(\mathfrak{U}_0) \sim 0$ mod $M - A_0$, say (with $i = 0$)

$$\beta\Omega_{n+i+1}(\mathfrak{U}_{i+1}) = \pi_{i+1}\rho_i\Omega_{n+i}(\mathfrak{U}_i) \text{ mod } M - A_i. \tag{2}$$

Now $\rho_1\Omega_{n+1}(\mathfrak{U}_1)$ is a cycle mod $M - A_0$ since its boundary mod $M - A_0$ is $\rho_0\rho_1(\pi_1\Omega_n) = \delta\sigma\pi_1\Omega_n$ which vanishes since $\delta\sigma = (1 - T)(1 + \cdots + T^{p-1}) = 1 - T^p = 0$. From the definition of the $\mathfrak{U}$'s we have $\pi_2\rho_1\Omega_{n+1}(\mathfrak{U}_1) \sim 0$ mod $M - A_1$. Hence there exists a relation (2) with $i = 1$. On repeating the construction we obtain relations (2) corresponding to $i = 0, \cdots, k$. Let

$$\begin{aligned} Y_{n+i} &= \pi_k\pi_{k-1}\cdots\pi_{i+1}\Omega_{n+i}(\mathfrak{U}_i) \qquad (i = 0, \cdots, k-1) \\ Y_{n+k} &= \Omega_{n+k}(\mathfrak{U}_k). \end{aligned} \tag{3}$$

Then from (2) we have

$$\beta Y_{n+i} = \rho_{i-1}Y_{n+i-1} \text{ mod } M - A_i \qquad (i = 1, \cdots, k). \tag{4}$$

Now ρ_iY_{n+i} is a $\mathfrak{U}_k$-cycle mod $M - A_i$, hence mod $M - A_k$. The particular cycle ρ_kY_{n+k} is null. For by lemma 3, this cycle lies in $\mathfrak{U}_k - (\mathfrak{U}_k)_0$. But this is impossible unless $\rho_kY_{n+k} = 0$ since $n + k = m + 1$ whereas by the properties of special coverings, the simplexes of $\mathfrak{U}_k - (\mathfrak{U}_k)_0$ are of dimension $\leqq n$. We conclude therefore that $\rho_kY_{n+k} \simeq 0$ mod $M - A_k$. Hence by successive applications of lemma 4, the relations (4) yield

$$\rho_iY_{n+i} \simeq 0 \text{ mod } M - A_k \qquad (i = k, \cdots, 0).$$

In particular $\rho_0Y_n \simeq 0$ mod $M - A_k$ and hence there exists a relation of the form

$$\beta\rho_0X_{n+1}(\mathfrak{U}_k) = \rho_0Y_n(\mathfrak{U}_k) \text{ mod } M - A_k.$$

Write

$$\beta X_{n+1}(\mathfrak{U}_k) = Y_n(\mathfrak{U}_k) - H(\mathfrak{U}_k). \tag{5}$$

On operating on both sides of (5) by ρ_0, we find that $\rho_0 H = \delta H = 0 \bmod M - A_k$. Hence by [3], section 3.11, there exist chains $\Gamma(\mathfrak{U}_k)$, $\Gamma^L(\mathfrak{U}_k)$, the latter in L, such that

$$H(\mathfrak{U}_k) = \sigma\Gamma(\mathfrak{U}_k) + \Gamma^L(\mathfrak{U}_k) \bmod M - A_k. \tag{6}$$

From (3) with $i = 0$, we have $Y_n \sim \Omega_n \bmod M - A_k$. Consequently from (5) and (6) we have

$$\Omega_n(\mathfrak{U}_k) \sim \sigma\Gamma(\mathfrak{U}_k) + \Gamma^L(\mathfrak{U}_k) \bmod M - A_k$$

which, since $\mathfrak{U}_k$ is an arbitrary special covering, establishes A) under the assumption that $m > n$. The proof for the case $m = n$ is not essentially different and we therefore omit further details.

B) In the cycle Δ of proposition A) the chains $\sigma\Gamma(\mathfrak{U}^T)$ and $\Gamma^L(\mathfrak{U}^T)$ are $\mathfrak{U}^T$-cycles mod $M - A_k$.

For in any case $\beta\Delta$ equals

$$\sigma\beta\Gamma(\mathfrak{U}^T) + \beta\Gamma^L(\mathfrak{U}^T) = 0 \bmod M - A_k. \tag{1}$$

Now $\beta\Gamma^L(\mathfrak{U}^T) \subset \mathfrak{U}_0^T$ whereas, by lemma 3, $\sigma\beta\Gamma(\mathfrak{U}^T) \subset \mathfrak{U}^T - \mathfrak{U}_0^T$. Hence the two chains in the left member of (1) have no common simplex and therefore vanish separately mod $M - A_k$.

C) Let d be a point in L^T. There exists for d over $\mathfrak{g}_p$ an n-dimensional fundamental pair (Δ, A) such that $A(d)$ is invariant and Δ is of the form

$$\{\sigma\Gamma(\mathfrak{U}^T) + \Gamma^L(\mathfrak{U}^T)\}, \qquad \Gamma^L(\mathfrak{U}^T) \subset L^T.$$

For by the hypothesis of uniform n-regularity there exists for d over $\mathfrak{g}_p$ an n-dimensional fundamental pair (Ω, D). Now since d is a point of L^T, any given neighborhood $C(d)$ will contain an invariant neighborhood of d. Such a neighborhood, for example, is the intersection of $C, TC, \cdots, T^{p-1}C$. Thus D can be replaced by an invariant neighborhood of d. Moreover, from the n-regularity of M over $\mathfrak{g}_p$, there exist neighborhoods $A_0, \cdots, A_k$ of d satisfying the conditions stated in A). The set $A = A_k$ forms together with the cycle Δ of A) the desired fundamental pair.

D) L^T is nowhere dense in M.

For, let $L = L^T$ and let L_0 be the maximal open subset of L. Assume, contrary to D) that $L_0 \neq 0$. Then $\bar{L}_0 - L \neq 0$, otherwise L_0 would be both open and closed, hence identical with M so that T would be the identity. Let d be a point in $\bar{L}_0 - L_0$ and let (Δ, A) be the fundamental pair for d of proposition C). On account of the n-regularity of M over $\mathfrak{g}_p$ there exists a neighborhood $A'(d) \subset A$ such that if C is a non-empty subset of A, n-cycles mod $M - A$ in $M - C$ are ~ 0 mod $M - A'$. Suppose in particular that $C = A' - A'L$. Then by the choice of d, $C \neq 0$. Moreover $L \subset M - C$ since $L(A' - A'L) = 0$.

Therefore $\Gamma^L(\mathfrak{U}^T) \subset M - C$ for every $\mathfrak{U}^T$. Let $\mathfrak{U}$ be an arbitrary but fixed $\mathfrak{U}^T$. If $\mathfrak{V} \in \{\mathfrak{U}^T\}$ is a suitably chosen refinement of $\mathfrak{U}^T$ and π an arbitrarily chosen projection $\mathfrak{V} \to \mathfrak{U}$, then $\pi\Gamma^L(\mathfrak{V})$ is the $\mathfrak{U}$-coordinate of a Čech cycle mod $M - A$ in $M - C$ (see proof of A)), hence is ~ 0 mod $M - A'$. (That $\Gamma^L(\mathfrak{V})$ is a cycle mod $M - A$ follows from proposition C)). Next take for C the non-empty set L_0A'. Then $\sigma\Gamma \subset M - C$ since in any case $\sigma\Gamma \subset M - L$ by lemma 3 and $M - L \subset M - L_0A' = M - C$. Hence as above, for a suitable special refinement $\mathfrak{V}_1$ of $\mathfrak{U}$ and projection $\pi_1 : \mathfrak{V}_1 \to \mathfrak{U}$, we have $\pi_1\sigma\Gamma(\mathfrak{V}_1) \sim 0$ mod $M - A'$. We may assume that $\mathfrak{V} = \mathfrak{V}_1$ for both coverings can be replaced by a common (special) refinement. If we then take $\pi = \pi_1$ we have

$$\pi(\sigma\Gamma(\mathfrak{V}) + \Gamma^L(\mathfrak{V})) \sim 0 \text{ mod } M - A'$$

and hence the $\mathfrak{U}$-coordinate of Δ is ~ 0 mod $M - A'$. Since $\mathfrak{U}$ is an arbitrary special covering, this implies that $\Delta \sim 0$ mod $M - A'$ which is impossible.

E) Let d be a point in N and let $F(D, d, n)$ be a uniform family of n-dimensional fundamental pairs (Ω^q, D) $(q = 2, 3, 5, \cdots)$ such that $D(d) \subset N$. There exists a neighborhood $B(d) \subset D$ and a covering $\mathfrak{V}$ of M such that if T is a periodic transformation of prime period p satisfying the relation: $T \ll \mathfrak{V}$ over N, there exist chains of the form $\sigma\Gamma(\mathfrak{U}^T)$ with coefficients in $\mathfrak{g}_p$ such that

$$\{\sigma\Gamma(\mathfrak{U}^T)\} \sim \Omega^p \text{ mod } M - B.$$

To prove this, choose a neighborhood $D'(d) \subset D$ such that for each prime q, Ω^q is a basis for n-cycles mod $M - D$ relative to homologies mod $M - D'$ (coefficients in $\mathfrak{g}_q$). On replacing D' by a smaller neighborhood of d if necessary, we may assume that $\bar{D}' \subset D$. Next, choose a neighborhood $D''(d)$ related to D' in the same way that D' is related to D. Now let $k = m - n + 1$ and choose neighborhoods $B_i(d)$, $B_i'(d)$ $(i = 0, \cdots, k)$ contained in D such that

$$B_i' \supset B_{i+1} \qquad (i = 0, \cdots, k - 1)$$

$$B_i \supset \bar{B}_i' \qquad (i = 0, \cdots, k)$$

and such that, for each q, cycles mod $M - B_i'$ over $\mathfrak{g}_q$ of dimension $> n$ are ~ 0 mod $M - B_{i+1}$ $(i = 0, \cdots, k - 1)$. By repeated applications of lemma 2 there exists a covering $\mathfrak{V}$ such that if $T \ll \mathfrak{V}$ over N, then $D \supset \sigma D'$ and $B_i \supset \sigma B_i'$, so that

$$D \supset \sigma D' \supset D' \supset D'' \supset B_0 \supset \sigma B_0' \supset B_0' \supset B_1 \supset \sigma B_1' \supset B_1' \supset \cdots.$$

In particular, if we write $G = \sigma D'$ and $A_i = \sigma B_i'$, we will have

$$D \supset G \supset A_0 \supset A_1 \supset \cdots \tag{1}$$

$$G \supset D' \supset D'' \supset A_0. \tag{2}$$

The sets $G, A_0, A_1, \cdots$ depend on T since σ does but the sets D', D'', B_i, B_i' do not. Consider $B_k(d)$. We can choose a neighborhood $B(d) \subset B_k$ such that

if C is a non-empty open subset of B and q a prime, then over $\mathfrak{g}_q$, n-cycles mod $M - B$ in $M - C$ are ~ 0 mod $M - B'$.

Now consider a definite T of prime period p such that $T \ll \mathfrak{B}$ over N. Since $G \subset D$, (Ω^p, G) is a fundamental pair over $\mathfrak{g}_p$. Moreover, it follows from (2) and the relation between D' and D'' that Ω^p is a basis over $\mathfrak{g}_p$ for n-cycles mod $M - G$ relative to homologies mod $M - A_0$. Finally, cycles mod $M - A_i$ over $\mathfrak{g}_p$ of dimension $> n$ are ~ 0 mod $M - A_{i+1}$; for, such cycles are cycles mod $M - B_i'$ since $B_i' \subset \sigma B_i' = A_i$ and as such, they are ~ 0 mod $M - B_{i+1}$, hence mod $M - A_{i+1}$ since $A_{i+1} = \sigma B_{i+1}' \subset B_{i+1}$. From these facts and the relations (1) it follows that the sets $G, A_0, \cdots, A_k$ satisfy the conditions in proposition A) and consequently there exists a cycle Δ mod $M - A_k$ over $\mathfrak{g}_p$ of the form

$$(3) \qquad \{\sigma\Gamma(\mathfrak{U}^T) + \Gamma^L(\mathfrak{U}^T)\}$$

such that $\Delta \sim \Omega^p$ mod $M - A_k$.

Let γ^T denote an arbitrary n-cycle mod $M - A_k$ in L^T, coefficients in $\mathfrak{g}_p$. γ^T is a cycle mod $M - B_k'$ since $B_k' \subset \sigma B_k' = A_k$. The open set $B_k' - L^T B_k'$ is non-empty since L^T is nowhere dense (proposition D)). Hence γ^T, being in L^T, hence in $M - (B_k' - L^T B_k')$, is ~ 0 mod $M - B$ by choice of B. In short, n-cycles mod $M - A_k$ (over $\mathfrak{g}_p$) in L^T are ~ 0 mod $M - B$. Now for each $\mathfrak{U}^T$, $\Gamma^L(\mathfrak{U}^T)$ is a cycle mod $M - A_k$ in L^T (proposition B)). If $\mathfrak{B}^T$ is a refinement of $\mathfrak{U}^T$ and π a projection $\mathfrak{B}^T \to \mathfrak{U}^T$, the $\mathfrak{U}^T$-coordinate of Δ can be replaced by the projection of its $\mathfrak{B}^T$-coordinate by π. But if $\mathfrak{B}^T$ is suitably chosen, $\pi\Gamma^L(\mathfrak{B}^T)$ will be the $\mathfrak{U}^T$-coordinate of a Čech cycle mod $M - A_k$ in L^T, coefficients in $\mathfrak{g}_p$, hence will be ~ 0 mod $M - B$. Hence we may assume in (3) that $\{\Gamma^L(\mathfrak{U}^T)\} \sim 0$ mod $M - B$. Hence $\Omega^p \sim \{\sigma\Gamma(\mathfrak{U}^T)\}$ mod $M - B$. This establishes E).

F) Let $\mathfrak{U}$ be a covering of M and let T be a periodic transformation such that $T \ll \mathfrak{U}$ over N. Let N^* be a non-empty open set such that $\bar{N}^* \subset N$. There exists a special refinement $\mathfrak{U}^T$ of $\mathfrak{U}$ and projection π: $\mathfrak{U}^T \to \mathfrak{U}$ such that $\pi T = \pi$ for all $\mathfrak{U}^T$-simplexes in N^*.

The component sets of a given $\mathfrak{U}^T$ fall into "cyclic families," the sets in a given family being images of each other under powers of T. Each family contains exactly p sets or a single (invariant) set. Let cU^T denote the cyclic family to which $U^T (\epsilon\, \mathfrak{U}^T)$ belongs. Let $\nu(\mathfrak{U}^T)$ denote the totality of cyclic families cU^T such that at least one member of cU^T meets N^*. To establish F) it will be sufficient to show that there exists a refinement $\mathfrak{U}^T$ of $\mathfrak{U}$ such that to each cyclic family cU^T in $\nu(\mathfrak{U}^T)$ there can be associated a set $U = U(cU^T)$ of $\mathfrak{U}$, containing all the sets in cU^T. For if π is a projection $\mathfrak{U}^T \to \mathfrak{U}$ carrying each set in cU^T (cU^T a member of $\nu(\mathfrak{U}^T)$) into $U(cU^T)$ and arbitrarily defined for the remaining sets of $\mathfrak{U}^T$, it is clear that $\pi T = \pi$ for $\mathfrak{U}^T$-complexes in N^*. To show that $\mathfrak{U}^T$ exists, let there be associated to each point x in N a set U_x of $\mathfrak{U}$ containing the orbit of x. This is possible since $T \ll \mathfrak{U}$ over N. Evidently we may associate to each x in N a neighborhood $V(x)$ such that the images of $V(x)$ under powers of T are contained in U_x. Let $V(x_1), \cdots, V(x_n)$ be a finite number of

these sets forming a covering of $\bar{N}^*$ and let Y be an open set such that $\bar{N}^* \subset Y$, $\bar{Y} \subset \Sigma V(x_i)$. Let $\mathfrak{X}$ be the totality of sets $V(x_i)$ together with the single set $M - \bar{Y}$. Evidently $\mathfrak{X}$ is a covering of M. Now the closed sets $\bar{N}^*$ and $M - Y$ are disjoint. Hence there can be chosen a $\mathfrak{U}^T$, refinement of $\mathfrak{X}$, such that sets of $\mathfrak{U}^T$ which meet $\bar{N}^*$ will not meet $M - Y$. Consider a cyclic family say cU_1^T in $\nu(\mathfrak{U}^T)$. At least one member of cU^T meets N^*. That member, however, can not meet $M - \bar{Y}$, hence it is contained in one of the sets $V(x_i)$ say $V(x_1)$. Hence its images—the members of cU_1^T—are all contained in U_{x_1}. If we define the function $U(cU)$ to be U_{x_1} for the cyclic family cU_1^T and similarly for the remaining cyclic families of $\nu(\mathfrak{U}^T)$, it will possess the desired properties.

G) There exists a covering $\mathfrak{A}$ of M such that no periodic T of prime period can satisfy the relation $T \ll \mathfrak{A}$ over N.

For let d be a point of N and let $F(D, d, n)$ be a uniform family of n-dimensional fundamental pairs (Ω^p, D). On replacing D by a smaller neighborhood of d if necessary, we may assume that $D \subset N$. Let $B(d)$, $\mathfrak{B}$ be the neighborhood and covering of proposition E) and let $H(d)$ be a neighborhood of d such that $\bar{H} \subset B$. Since $F(D, d, n)$ uniformly satisfies condition (1) in the definition of fundamental pair (section 4) there exists a covering $\mathfrak{U}$ such that for arbitrary prime p,

$$\Omega^p(\mathfrak{U}) \nsim 0 \bmod M - H. \tag{1}$$

Let $\mathfrak{B}$ be a covering such that if T is a periodic transformation with $T \ll \mathfrak{B}$ over N, then $\sigma\bar{H} \subset B$ (lemma 2). Let $\mathfrak{A}$ be a common refinement of $\mathfrak{B}$, $\mathfrak{U}$, $\mathfrak{B}$. We shall show that $\mathfrak{A}$ has the required property. Suppose there exists a T of prime period p such that $T \ll \mathfrak{A}$ over N. Let $N^* = \sigma H$. Then $\bar{N}^* \subset B \subset N$. By proposition F) there exists a refinement $\mathfrak{U}^T$ of $\mathfrak{U}$ and projection π: $\mathfrak{U}^T \to \mathfrak{U}$ such that $\pi T = \pi$ for $\mathfrak{U}^T$-simplexes in N^*. Consider the relative cycle $\{\sigma\Gamma(\mathfrak{U}^T)\}$ of proposition E). Let $\Lambda(\mathfrak{U}^T)$ be the subchain of $\Gamma(\mathfrak{U}^T)$ consisting of these simplexes which are in N^*. Then $\Gamma(\mathfrak{U}^T) - \Lambda(\mathfrak{U}^T) \subset M - N^*$. Since N^* is invariant under T, we have also

$$\sigma\Gamma(\mathfrak{U}^T) - \sigma\Lambda(\mathfrak{U}^T) \subset M - N^*.$$

This relation holds if the left member is replaced by its image under π. But since the chain $\sigma\Lambda(\mathfrak{U}^T)$ is in N^*, its image under π is $p\Lambda(\mathfrak{U}^T) = 0 \bmod p$. Hence $\pi\sigma\Gamma(\mathfrak{U}^T) \sim 0 \bmod M - N^*$. Since $\Omega^p(\mathfrak{U}^T) \sim \sigma\Gamma(\mathfrak{U}^T) \bmod M - B$, hence mod $M - H$, we have $\pi\Omega^p(\mathfrak{U}^T) \sim 0 \bmod M - H$. By definition of Čech cycle, $\pi\Omega^p(\mathfrak{U}^T) \sim \Omega^p(\mathfrak{U}) \bmod M - H$ so that $\Omega^p(\mathfrak{U}) \sim 0 \bmod M - H$ which contradicts (1). This concludes the proof of theorem II under the hypothesis that M is bicompact.

7. We conclude with a word concerning the locally bicompact case. In the first place, an examination of the preceding proof shows that in the bicompact case, it would have been sufficient to assume merely that M is uniformly n-regular over N instead of the whole of M. Now suppose that M, N satisfy the

conditions of the theorem but that M is not bicompact. M can be converted into a bicompact space M_1 by the addition of a single point[8] say ∞. In M_1, N has for its counterpart an open set N_1 not containing ∞. If M is uniformly n-regular, then M_1 is at least uniformly n-regular over N_1. Hence there is a covering $\mathfrak{A}_1$ of M_1 such that no periodic transformation T_1 operating in M_1 can satisfy the relation $T_1 \ll \mathfrak{A}_1$ over N_1. Now every periodic T operating in M induces a periodic T_1, of same period, operating in M_1 and leaving ∞ fixed. If $\mathfrak{A}$ is the covering of M obtained from $\mathfrak{A}_1$ by suppressing the point ∞, it is easy to see that the relation: $T \ll \mathfrak{A}$ over N would imply $T \ll \mathfrak{A}_1$ over N_1. Hence $\mathfrak{A}$ is the required covering of M.

THE INSTITUTE FOR ADVANCED STUDY AND BARNARD COLLEGE,
COLUMBIA UNIVERSITY

REFERENCES

(1) M. H. A. Newman, A theorem on periodic transformations of spaces, Quart. Journal 2 (1931) pp. 1–8.
(2) E. Čech, Théorie générale de l'homologie dans un espace quelconque, Fund. Math. 18 (1932) pp. 149–183.
(3) P. A. Smith, Transformations of finite period, Annals of Math. 39 (1938) pp. 127–164.
(4) P. A. Smith, Transformations of finite period. II, Annals of Math. 40 (1939) 690–711.

[8] We may define as neighborhood of ∞ the complement of any bicompact set in M, plus ∞ itself. With this topology M_1, like M, is a connected finite-dimensional space in which open sets are countable sums of closed sets. Moreover, M_1 will be uniformly n-regular except possibly at the point ∞.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

CONTINUOUS MAPPINGS OF INFINITE POLYHEDRA

BY SAMUEL EILENBERG

(Received March 20, 1940)

This paper is concerned with some special cases of the general problem of determining the homotopy classes of continuous mappings of a polyhedron K into a space Y. The well known theorem of H. Hopf solves the problem when K is a finite n-dimensional polyhedron and $Y = S^n$ is the n-sphere. The solution is given in terms of homology. In a recent paper[1] the author has generalized Hopf's result simultaneously in the following three directions:

(a) The hypothesis that K is finite is removed by admitting infinite chains and letting cohomologies replace homologies.

(b) The hypothesis that K is n-dimensional is replaced by a hypothesis concerning the vanishing of certain higher-dimensional cohomology groups.

(c) S^n is replaced by a more general space Y.

Although the use of cohomologies in this problem seems quite natural and justified, the results become more intuitive and easier to apply in the cases when the theorem has a homological interpretation. Such an interpretation is easily available when K is finite but does not seem to be possible generally. Section **2** of this paper gives such an interpretation for a large class of infinite polyhedra. Various applications are given in sections **3**–**10**. The results of **CM** which are being used here are collected in section **1**.

1. Let K be a finite or infinite cell-complex, i.e. a locally finite polyhedron with a given cell decomposition. K^n will denote the sub-complex of K consisting of all cells of dimension $\leqq n$. Given an abelian group G we shall consider infinite chains in K with coefficients in G and denote the resulting n-dimensional cohomology group by ${}^n\boldsymbol{H}_G(K)$.

Let Y be an arcwise connected topological space and let $y_0 \in Y$. We shall denote by $\pi_i(Y)$ the i^{th} homotopy group of Y with respect to y_0 as origin.[2] We shall assume that $\pi_i(Y) = 0$ for $i < n$. If $n = 1$ we replace this condition by the condition that $\pi_1(Y)$ (= the fundamental group of Y) is abelian.

The set of all continuous mappings $f(K) \subset Y$ will be denoted by Y^K. Given $f_0, f_1 \in Y^K$ we use the notation $f_0 \simeq f_1$ to denote that f_0 and f_1 are homotopic.

(1.1) *Given $f \in Y^K$ there is an $f' \in Y^K$ such that $f \simeq f'$ and $f'(K^{n-1}) = y_0$* [**CM**, p. 241]

[1] Cohomology and continuous mappings, these *Annals*, vol. 41 (1940), pp. 231–251. We shall refer to this paper as to **CM**.

[2] This group was introduced by W. Hurewicz, *Proc. Akad. Amsterdam* 38 (1935), p. 113. The definition given in **CM**, p. 235, seems to suit our purposes best.

Given $f \in Y^K$ such that $f(K^{n-1}) = y_0$ we construct an n-dimensional chain $d^n(f)$ in K with coefficients from $\pi_n(Y)$ as follows:

Let σ^n be an oriented n-cell of K. Since the boundary of σ^n is carried by f into the point y_0, the mapping $f(\sigma^n) \subset Y$ determines uniquely an element $d(f, \sigma^n)$ of $\pi_n(Y)$. We define

$$d^n(f) = \sum_i d(f, \sigma_i^n)\sigma_i^n$$

We have proved that

(1.2) *$d^n(f)$ is a cocycle in K.* [**CM**, p. 240, Hom. Th. 1]

(1.3) *If $f_0, f_1 \in Y^K$, $f_0(K^{n-1}) = f_1(K^{n-1}) = y_0$ and $f_0 \simeq f_1$ then $d^n(f_0)$ and $d^n(f_1)$ are cohomologous in K.* [**CM**, p. 242, Hom. Th. 3]

Given an arbitrary $f \in Y^K$ there is by (1.1) a mapping $f' \in Y^K$ such that $f \simeq f'$ and $f'(K^{n-1}) = y_0$. From (1.3) it follows that the cohomology class of the cocycle $d^n(f')$ is independent of the choice of f'. We denote this cohomology class by $\boldsymbol{d}^n(f)$. It follows from (1.3) that

(1.4) *If $f_0, f_1 \in Y^K$ and $f_0 \simeq f_1$ then $\boldsymbol{d}^n(f_0) = \boldsymbol{d}^n(f_1)$.*

It follows that for each homotopy class Φ of Y^K the cohomology class $\boldsymbol{d}^n(\Phi)$ is uniquely defined.

The following theorem includes all the further results of **CM** which will be needed in the sequel [**CM**, p. 243].

THEOREM 1. *Let K be a finite or infinite cell-complex and Y an arcwise connected topological space such that $\pi_i(Y) = 0$ for $i < n$. If $n = 1$ this condition is replaced by the condition that Y is i-simple*[3] *for $i = 1, 2, \cdots$, dim K.*

(1.5) *If ${}^{i+1}\boldsymbol{H}_{\pi_i(Y)}(K) = 0$ for $i = n + 1, n + 2, \cdots$ then given a cocycle d^n of K with coefficients in $\pi_n(Y)$ there is a mapping $f \in Y^K$ such that $f(K^{n-1}) = y_0$ and $d^n(f) = d^n$.*

(1.6) *If ${}^{i}\boldsymbol{H}_{\pi_i(Y)}(K) = 0$ for $i = n + 1, n + 2, \cdots$ then, given $f_0, f_1 \in Y^K$ such that $f_0(K^{n-1}) = f_1(K^{n-1}) = y_0$, we have $f_0 \simeq f_1$ if and only if $d^n(f_0)$ and $d^n(f_1)$ are cohomologous in K.*

(1.7) *If ${}^{i}\boldsymbol{H}_{\pi_i(Y)}(K) = {}^{i+1}\boldsymbol{H}_{\pi_i(Y)}(K) = 0$ for $i = n + 1, n + 2, \cdots$ then the homotopy classes Φ of Y^K are in a (1-1)-correspondence with the elements of the cohomology group ${}^{n}\boldsymbol{H}_{\pi_n(Y)}(K)$. The correspondence is determined by the operation $\boldsymbol{d}^n(\Phi)$.*

2. Let ${}^n\mathcal{H}^I(K)$ be the n^{th} homology group of K obtained using finite chains, and ${}^n\mathcal{H}^I(Y)$ the n^{th} homology group of Y obtained using finite singular chains. In both cases the group I of integers is taken as the coefficient group.

It is well known that every $f \in Y^K$ induces a homomorphic mapping h_f^n of ${}^n\mathcal{H}^I(K)$ into a subgroup of ${}^n\mathcal{H}^I(Y)$ and also that if $f_0, f_1 \in Y^K$ and $f_0 \simeq f_1$ then $h_{f_0}^n = h_{f_1}^n$. It follows that for every homotopy class Φ of Y^K the homomorphism h_Φ^n is uniquely defined.

[3] See **CM**, p. 235. The condition was introduced by the author (*Fund. Math.* 32 (1939), pp. 167–175) and it concerns the mutual behavior of $\pi_1(Y)$ and $\pi_i(Y)$. In particular Y is i-simple if one of these groups vanishes. Y is 1-simple if and only if $\pi_1(Y)$ is abelian.

Let us assume that K is connected, Y is arcwise connected and that $\pi_i(K) = \pi_i(Y) = 0$ for $i < n$. If $n = 1$ the last condition will be replaced by the condition that $\pi_1(K)$ and $\pi_1(Y)$ are abelian. In this case it was proved by W. Hurewicz that the groups $\pi_n(K)$ and ${}^n\mathcal{H}^I(K)$ can be considered identical.[4] Similarly for $\pi_n(Y)$ and ${}^n\mathcal{H}^I(Y)$. The homomorphism h_f^n induced by a mapping $f \in Y^K$ becomes then a homomorphic mapping of $\pi_n(K)$ into a subgroup of $\pi_n(Y)$.

THEOREM 2. *Let K be a connected finite or infinite cell-complex and Y an arcwise connected topological space such that $\pi_i(K) = \pi_i(Y) = 0$ for $i < n$. If $n = 1$ this last condition is replaced by the condition that K is 1-simple and Y is i-simple for $i = 1, 2, \cdots, \dim K$.*

(2.1) *If ${}^{i+1}H_{\pi_i(Y)}(K) = 0$ for $i = n + 1, n + 2, \cdots$ then given a homomorphism h mapping $\pi_n(K)[= {}^n\mathcal{H}^I(K)]$ into a subgroup of $\pi_n(Y)[= {}^n\mathcal{H}^I(Y)]$ there is an $f \in Y^K$ such that $h_f^n = h$.*

(2.2) *If ${}^iH_{\pi_i(Y)}(K) = 0$ for $i = n + 1, n + 2, \cdots$ then given $f_0, f_1 \in Y^K$ we have $h_{f_0}^n = h_{f_1}^n$ if and only if $f_0 \simeq f_1$.*

(2.3) *If ${}^iH_{\pi_i(Y)}(K) = {}^{i+1}H_{\pi_i(Y)}(K) = 0$ for $i = n + 1, n + 2, \cdots$ then the homotopy classes Φ of Y^K are in a (1-1)-correspondence with the homomorphic mappings of $\pi_n(K)[= {}^n\mathcal{H}^I(K)]$ into subgroups of $\pi_n(Y)[= {}^n\mathcal{H}^I(Y)]$. The correspondence is determined by the operation h_Φ^n.*

PROOF. Let $k_0 \in K^{n-1}$. Consider the mapping $1 \in K^K$ such that $1(x) = x$ for every $x \in K$. According to (1.1) there is a $g \in K^K$ such that $g \simeq 1$ and $g(K^{n-1}) = k_0$. According to (1.2) this leads to a cocycle

$$d^n(g) = \sum_i d(g, \sigma_i^n)\sigma_i^n$$

in K with coefficients in $\pi_n(K)$.

Suppose now that conditions of (2.1) are satisfied and let h be a homomorphic mapping of $\pi_n(K)$ into a subgroup of $\pi_n(Y)$. Consider the cocycle

$$d^n = \sum_i h[d(g, \sigma_i^n)]\sigma_i^n$$

in K with coefficients in $\pi_n(Y)$. By (1.5) of Th. 1 there is an $f \in Y^K$ such that $f(K^{n-1}) = y_0$ and $d^n(f) = d^n$. We are going to prove that $h_f^n = h$.

Since $\pi_n(K)$ and ${}^n\mathcal{H}^I(K)$ are identical, every element $a \in \pi_n(K)$ can be represented as a finite cycle

$$\sum_i \alpha_i\sigma_i^n$$

where α_i are integers and only a finite number of them are $\neq 0$. Using the definition of $d(g, \sigma_i^n)$ and of $d(f, \sigma_i^n)$ it follows that

$$h_g^n(a) = \sum_i \alpha_i d(g, \sigma_i^n), \qquad h_f^n(a) = \sum_i \alpha_i d(f, \sigma_i^n)$$

[4] *Proc. Akad. Amsterdam* 38 (1935), p. 521.

But since $d(f, \sigma_i^n) = h[d(g, \sigma_i^n)]$ it follows that

$$h_f^n(a) = h[h_g^n(a)].$$

Since $g \simeq 1$ we have $h_g^n(a) = a$; therefore $h_f^n(a) = h(a)$ and $h_f^n = h$. This proves (2.1).

Let $f_0, f_1 \in Y^K$ and let $h_{f_0}^n = h_{f_1}^n$. Without restricting the generality we may assume that $f_0(k_0) = f_1(k_0) = y_0$. Consider the mapping $f_j g \in Y^K$ for $j = 0, 1$. Since $f_j g(K^{n-1}) = y_0$ we may consider the cocycles $d^n(f_j g)$ in K with coefficients in $\pi_n(Y)$. Given any n-cell σ_i^n of K we get, according to the meaning of h_j^n,

$$d(f_j g, \sigma_i^n) = h_j^n[d(g, \sigma_i^n)].$$

This implies $d(f_0 g, \sigma_i^n) = d(f_1 g, \sigma_i^n)$, since $h_0^n = h_1^n$. Therefore $d^n(f_0 g) = d^n(f_1 g)$.

Now, if the hypothesis of (2.2) is satisfied, then by (1.6) of Th. 1 we have $f_0 g \simeq f_1 g$. Since $f_0 g \simeq f_0$ and $f_1 g \simeq f_1$ this gives $f_0 \simeq f_1$. Hence (2.2) is proved.

(2.3) follows immediately from (2.1) and (2.2).

3. Let G be an abelian group. Given a finite cycle

$$a^n = \sum_i \alpha_i \sigma_i^n$$

in K with integer coefficients and an arbitrary cocycle

$$d^n = \sum_i \beta_i \sigma_i^n$$

in K with coefficients in G, we define an element $a^n d^n$ of G by taking

$$a^n d^n = \sum_i \alpha_i \beta_i$$

It is easy to verify that if a^n is the boundary of a finite $(n+1)$-chain or if d^n is the coboundary of a $(n-1)$-chain then $a^n d^n = 0$. Consequently given two elements $\mathbf{a}^n \in {}^n\mathcal{H}^I(K)$ and $\mathbf{d}^n \in {}^nH_G(K)$ the multiplication $\mathbf{a}^n\mathbf{d}^n$ is a uniquely defined element of G. For a fixed $\mathbf{a}^n$ it gives a homomorphic mapping of ${}^n\boldsymbol{H}_G(K)$ into a subgroup of G. Similarly for fixed $\mathbf{d}^n$ it gives a homomorphic mapping of ${}^n\mathcal{H}^I(K)$ into a subgroup of G.

Now, let Y be an arcwise connected topological space such that $\pi_i(Y) = 0$ for $i < n$. If $n = 1$ we assume that Y is 1-simple instead. We choose G to be the group $\pi_n(Y)$ which can be considered identical with ${}^n\mathcal{H}^I(Y)$.

Let $f \in Y^K$ and let $f(K^{n-1}) = y_0$. We shall consider the cocycle

$$d^n(f) = \sum_i d(f, \sigma_i^n)\sigma_i^n$$

and its relation to the homomorphic mapping h_f^n of ${}^n\mathcal{H}^I(K)$ into a subgroup of ${}^n\mathcal{H}^I(Y)$ induced by f.

Given a finite cycle

$$a^n = \sum_i \alpha_i \sigma_i^n$$

in K with integer coefficients, we clearly have

$$h_f^n(a^n) = \sum_i \alpha_i\, d(f, \sigma_i^n)$$

and therefore

$$a^n d^n(f) = h_f^n(a^n).$$

Using (1.1) we find

(3.1) *Given any mapping $f \in Y^K$ we have $\boldsymbol{a}^n \boldsymbol{d}^n(f) = h_f^n(\boldsymbol{a}^n)$ for all $\boldsymbol{a}^n \in {}^n\mathcal{H}^I(K)$.*

Our hypothesis concerning Y is satisfied if $Y = S^n$ is the n-dimensional spherical manifold. The group $\pi_n(Y) = {}^n\mathcal{H}^I(Y)$ is then isomorphic to I. Using Th. 2 and (3.1) we prove

THEOREM 3. *Let K be a finite or infinite connected cell-complex such that $\pi_i(K) = 0$ for $i < n$. If $n = 1$ we replace this condition by the condition that $\pi_1(K)$ is abelian.*

(3.2) *If ${}^{i+1}\boldsymbol{H}_{\pi_i(S^n)}(K) = 0$ for $i = n + 1, n + 2, \cdots$ then for every homomorphic mapping h of the group ${}^n\mathcal{H}^I(K)$ into a subgroup of I there is a cohomology class $d^n \in {}^n\boldsymbol{H}_I(K)$ such that $\boldsymbol{a}^n\boldsymbol{d}^n = h(\boldsymbol{a}^n)$ for every $\boldsymbol{a}^n\,{}^n\mathcal{H}^I(K)$.*

(3.3) *If ${}^i\boldsymbol{H}_{\pi_i(S^n)}(K) = {}^{i+1}\boldsymbol{H}_{\pi_i(S^n)}(K) = 0$ for $i = n + 1, n + 2, \cdots$ then given $\boldsymbol{d}^n \in {}^n\boldsymbol{H}_I(K)$ such that $\boldsymbol{a}^n\boldsymbol{d}^n = 0$ for all $\boldsymbol{a}^n \in {}^n\mathcal{H}^I(K)$ we have $\boldsymbol{d}^n = 0$.*

It follows from (3.2) and (3.3) that if the hypothesis of (3.3) is satisfied then the homomorphisms h mapping the group ${}^n\mathcal{H}^I(K)$ into subgroups of I and the elements $\boldsymbol{d}^n$ of ${}^n\boldsymbol{H}_I(K)$ are in a (1-1)-correspondence. h and $\boldsymbol{d}^n$ correspond to each other if and only if

$$\boldsymbol{a}^n\boldsymbol{d}^n = h(\boldsymbol{a}^n) \text{ for all } \boldsymbol{a}^n \in {}^n\mathcal{H}^I(K).$$

PROOF OF TH. 3. If the hypothesis of (3.2) is satisfied then given the homomorphism h there is by (2.1) of Th. 2 a mapping $f \in S^{n^K}$ such that $h_f^n = h$. By (3.1) we then have $\boldsymbol{a}^n\boldsymbol{d}^n(f) = h(\boldsymbol{a}^n)$ for every $\boldsymbol{a}^n \in {}^n\mathcal{H}^I(K)$. This proves (3.2).

If the hypothesis of (3.3) is satisfied then by (2.1) of Th. 2 there is for every $\boldsymbol{d}^n$ a mapping $f \in Y^K$ such that $\boldsymbol{d}^n(f) = \boldsymbol{d}^n$. If now $\boldsymbol{a}^n\boldsymbol{d}^n = 0$ for all $\boldsymbol{a}^n \in {}^n\mathcal{H}^I(K)$ then it follows from (3.1) that $h_f^n = 0$. By (2.2) of Th. 2 we then have $f \simeq f_0$ where $f_0(K) = y_0$. This implies, by (1.6) of Th. 1, that $\boldsymbol{d}^n(f_0) = \boldsymbol{d}^n(f)$. Therefore $\boldsymbol{d}^n = 0$ and (3.3) is proved.

The following example will show that in Th. 3 the hypothesis that $\pi_i(K) = 0$ for $i < n$ cannot generally be removed if $n > 1$. Let Σ be one of D. van Dantzig's solenoids[5] imbedded in S^3. Let K^3 be a subdivision of the open and connected set $S^3 - \Sigma$ into an infinite cell-complex. It is clear that ${}^i\boldsymbol{H}_G(K^3) = 0$ for $i = 3, 4, \cdots$ and any abelian group G. It has been recently proved by N. E. Steenrod[6] that the 1^{st} homology group of K^3, constructed using infinite cycles and integer coefficients, is not enumerable. Since K^3 is a manifold this implies immediately that ${}^2\boldsymbol{H}_I(K^3)$ is not enumerable. On the other hand since Σ is connected it follows from duality theorems that ${}^2\mathcal{H}^I(K^3) = 0$. Hence (3.3)

---

[5] *Fund. Math.* 15 (1930), pp. 102–125.

[6] These *Annals*, vol. 41 (1940), pp. 833–851.

does not hold for $K = K^3$ and $n = 2$. Naturally, the condition $\pi_1(K^3) = 0$ is not satisfied.

4. Two topological spaces X and Y are said to have the same *homotopy type*[7] if there are two mappings $f \in Y^X$ and $g \in X^Y$ such that $gf \in X^X$ is homotopic to the mapping $1 \in X^X$ defined by the condition $1(x) = x$ for every $x \in X$, and similarly $fg \in Y^Y$ is homotopic to $1 \in Y^Y$.

THEOREM 4. *Two finite or infinite connected cell-complexes K_1 and K_2 which satisfy the following set of conditions have the same homotopy type:*

(4.1) $\pi_i(K_1) = \pi_i(K_2) = 0$ *for* $i < n$,

(4.2) $\pi_n(K_1)$ *and* $\pi_n(K_2)$ *are isomorphic,*

(4.3) ${}^{i+1}H_{\pi_i(K_2)}(K_1) = {}^{i+1}H_{\pi_i(K_1)}(K_2) = 0$ *for* $i = n + 1, n + 2, \cdots$,

(4.4) ${}^{i}H_{\pi_i(K_1)}(K_1) = {}^{i}H_{\pi_i(K_2)}(K_2) = 0$ *for* $i = n + 1, n + 2, \cdots$.

If $n = 1$ *we replace* (4.1) *by the following*

$(4.1)_1$ K_1 *and* K_2 *are i-simple for* $i = 1, 2, \cdots, k$, *where* $k = \max(\dim K_1, \dim K_2)$.

PROOF. Let h be an isomorphic mapping of $\pi_n(K_1)$ on $\pi_n(K_2)$ and let h^* be the inverse isomorphism. Applying (2.1) of Th. 2 to $K_2^{K_1}$ we obtain a mapping $f \in K_2^{K_1}$ such that $h_f^n = h$. Similarly we obtain a mapping $g \in K_1^{K_2}$ such that $h_g^n = h^*$. Consequently $h_{gf}^n = h^*h$. It follows that $h_{gf}^n = h_1^n$, where $1 \in K_1^{K_1}$ is defined by $1(x) = x$ for all $x \in K_1$. Applying (2.2) of Th. 2 we find that $gf \simeq 1$. Similarly $fg \simeq 1$ where $1 \in K_2^{K_2}$. This proves that K_1 and K_2 have the same homotopy type.

5. In the following we are going to characterize complexes which have the same homotopy type as S^n. The cases $n > 1$ and $n = 1$ will be treated separately.

THEOREM 5. *A finite or infinite connected cell-complex K has the homotopy type of S^n $(n > 1)$ if and only if*

(5.1) $$\pi_1(K) = 0,$$

(5.2) $${}^{i}\mathcal{H}^{I}(K) = 0 \textit{ for } 1 < i < n,$$

(5.3) $${}^{n}\mathcal{H}^{I}(K) \textit{ is cyclic infinite},$$

(5.4) $${}^{i+1}H_{\pi_i(S^n)}(K) = 0 \text{ for } i = n + 1, n + 2, \cdots,$$

(5.5) $${}^{i}H_{\pi_i(K)}(K) = 0 \text{ for } i = n + 1, n + 2, \cdots.$$

THEOREM 5_1. *A finite or infinite connected cell-complex K has the homotopy type of S^1 if and only if*

(5.6) $$\pi_1(K) \textit{ is cyclic infinite},$$

(5.7) $$K \textit{ is i-simple for } i = 2, 3, \cdots, \dim K,$$

(5.8) $${}^{i}H_{\pi_i(K)}(K) = 0 \textit{ for } i = 2, 3, \cdots.$$

[7] W. Hurewicz, *Proc. Akad. Amsterdam* 39 (1936), p. 124.

PROOFS. Sufficiency. $n > 1$. By a theorem of Hurewicz[4] it follows from (5.1) and (5.2) that $\pi_i(K) = 0$ for $i < n$. Therefore

$$\pi_i(K) = \pi_i(S^n) = 0 \text{ for } i < n. \tag{5.9}$$

From the same theorem it follows that $\pi_n(K)$ and ${}^n\mathcal{H}^I(K)$ are isomorphic. Therefore

$$\pi_n(K) \text{ and } \pi_n(S^n) \text{ are isomorphic.} \tag{5.10}$$

Since S^n is n-dimensional therefore

$${}^iH_G(S^n) = 0 \text{ for } i = n+1, n+2, \cdots \text{ and any } G. \tag{5.11}$$

From (5.4), (5.5), (5.9), (5.10) and (5.11) it follows that the conditions of Th. 4 are satisfied. Therefore K and S^n have the same homotopy type.

$n = 1$. Since $\pi_1(K)$ is cyclic infinite therefore we see that

$$\pi_1(K) \text{ and } \pi_1(S^1) \text{ are isomorphic.} \tag{5.12}$$

Since S^1 is i-simple for every i and $\pi_1(K)$ is abelian it follows from (5.7) that

$$K \text{ and } S^1 \text{ are } i\text{-simple for } i = 1, 2, \cdots, \dim K. \tag{5.13}$$

Since S^1 is 1-dimensional and $\pi_i(S^1) = 0$ for $i > 1$ it follows that

$${}^iH_G(S^1) = 0 \text{ for } i = 2, 3, \cdots \text{ and any } G, \tag{5.14}$$

$${}^{i+1}H_{\pi_i(S^1)}(K) = 0 \text{ for } i = 2, 3, \cdots . \tag{5.15}$$

From (5.8) and (5.12)–(5.15) it follows that the conditions of Th. 4 are satisfied and therefore K and S^1 have the same homotopy type.

The necessity follows from the following two lemmas:

(5.16) *If X and Y are arcwise connected and have the same homotopy type then $\pi_i(X)$ is isomorphic with $\pi_i(Y)$, and X is i-simple if and only if Y is.*

(5.17) *If the finite or infinite connected cell-complexes K_1 and K_2 have the same homotopy type then ${}^iH_G(K_1)$ is isomorphic to ${}^iH_G(K_2)$, and ${}^i\mathcal{H}^G(K_1)$ is isomorphic to ${}^i\mathcal{H}^G(K_2)$.*

The proofs are left out.

Since $\pi_i(K) = 0$ implies that K is i-simple it follows that if $\pi_1(K)$ is cyclic infinite and $\pi_i(K) = 0$ for $i > 1$, then the conditions of Th. 5_1 are satisfied and K has the homotopy type of S^1. This is a theorem of Hurewicz.[8]

6. Theorem 5 leads to the following characterization of the finite complexes which have the homotopy type of S^n.

THEOREM 6. *A finite polyhedron K has the homotopy type of S^n if and only if K has the same fundamental group and the same homology groups (integer*

[8] *Proc. Akad. Amsterdam* 39 (1936), p. 221.

[9] If k is n-dimensional, this is a theorem of Hurewicz, loc. cit. In the general case it also follows from some unpublished results of Hurewicz.

coefficients) as S^n.[9] *If* $n = 1$ *there is an additional condition that* K *is to be* i*-simple for* $i = 2, 3, \cdots, \dim K$.

PROOF. The necessity follows from (5.16) and (5.17). If K has the same integer coefficient homology groups as S^n then since K is finite it follows that ${}^i\boldsymbol{H}_G(K)$ is isomorphic with ${}^i\boldsymbol{H}_G(S)$ for any i and any coefficient group G. This shows that if the conditions of Th. 6 are satisfied then the conditions of Th. 5 (or Th. 5_1) are satisfied too, and consequently K has the homotopy type of S^n.

There is no example known to prove that the additional condition in the case $n = 1$ is really necessary in Th. 6.

7. Theorem 4 will be used again to prove the following theorem concerning homotopy groups.

THEOREM 7. *Let* K^{n+1} *be an at most* $(n + 1)$*-dimensional* $(n > 1)$ *finite or infinite connected cell-complex such that* $\pi_i(K^{n+1}) = 0$ *for* $i < n$. *If the group* $\pi_n(K^{n+1})$ *is a free group with a finite number* $k > 0$ *of generators then* $\pi_{n+1}(K^{n+1}) \neq 0$.

PROOF. Suppose that $\pi_{n+1}(K^{n+1}) = 0$, then

$$(7.1) \qquad {}^{n+1}\boldsymbol{H}_{\pi_{n+1}(K^{n+1})}(K^{n+1}) = 0.$$

Since K^{n+1} is at most $(n + 1)$-dimensional we have

$$(7.2) \qquad {}^i\boldsymbol{H}_G(K^{n+1}) = 0 \text{ for } i > n + 1 \text{ and any } G.$$

Let K_1^n be a complex obtained by taking k copies of S^n disjoint except for one point which is common for all of them. It is clear that

$$(7.3) \qquad \pi_i(K^{n+1}) = \pi_i(K_1^n) = 0 \text{ for } i < n,$$

$$(7.4) \qquad \pi_n(K^{n+1}) \text{ and } \pi_n(K_1^n) \text{ are isomorphic,}$$

$$(7.5) \qquad {}^i\boldsymbol{H}_G(K_1^n) = 0 \text{ for } i > n \text{ and any } G.$$

It follows from (7.1)-(7.5) that Th. 4 can be applied to demonstrate that K^{n+1} and K_1^n have the same homotopy type. Since $\pi_{n+1}(S^n) \neq 0$ for $n > 1$[10] it follows that $\pi_{n+1}(K_1^n) \neq 0$ and from (5.16) that $\pi_{n+1}(K^{n+1}) \neq 0$.

Th. 7 can also be proved in the case when $\pi_n(K^{n+1})$ is the unrestricted direct product of $\aleph_0$ cyclic infinite groups.

8. In the following we shall discuss the situation arising when S^{r-n-1} is topologically imbedded into S^r.

THEOREM 8. *Let* S_1^{r-n-1} *be a homeomorphic image of* S^{r-n-1} *contained in* S^r *where* $r > n > 0$.

If $n > 1$ *then* $S^r - S_1^{r-n-1}$ *has the homotopy type of* S^n *if and only if* $\pi_1(S^r - S_1^{r-n-1}) = 0$.

[10] L. Pontrjagin, *C. R. Acad. Sci. URSS* 19 (1938), p. 147; H. Freudenthal, *Comp. Math.* 5 (1937), p. 301.

If $n = 1$ then $S^r - S_1^{r-2}$ has the homotopy type of S^1 if and only if the group $\pi_1(S^r - S_1^{r-2})$ is cyclic infinite and $S^r - S_1^{r-2}$ is i-simple for $i = 2, 3, \cdots r$.

PROOF. The necessity of the conditions follows from (5.16). In order to prove that they are also sufficient note that it follows from duality theorems that

(8.1) ${}^{i}\mathcal{H}^{G}(S^r - S_1^{r-n-1})$ and ${}^{i}\mathcal{H}^{G}(S^n)$ are isomorphic for $i = 0, 1, \cdots$ and any G.

Since $S^r - S_1^{r-n-1}$ is a manifold the group ${}^{i}H_G(S^r - S_1^{r-n-1})$ is isomorphic with the $(r - i)$-th homology group of $S^r - S_1^{r-n-1}$ obtained using infinite chains. Since S_1^{r-n-1} is locally connected in all dimensions it follows from a recent theorem of Steenrod[6] that this group is isomorphic with ${}^{r-i-1}\mathcal{H}^{G}(S_1^{r-n-1})$. Therefore

(8.2) ${}^{i}H_G(S^r - S_1^{r-n-1}) = 0$ for $i = n + 1, n + 2, \cdots$ and any G.

Now, (8.1) and (8.2) and the conditions of Th. 8 permit us to apply Th. 5 (or Th. 5_1) and prove that $S^r - S_1^{r-n-1}$ has the homotopy type of S^n.

9. A set $Y \subset X$ is called a *retract* of X if there is a mapping $r \in X^X$ (called a *retraction*) such that $r(X) = Y$ and $r(y) = y$ for all $y \in Y$. If $r \simeq 1$, where $1 \in X^X$ is defined by $1(x) = x$ for all $x \in X$, then Y is called a *deformation retract* of X.

We shall consider in S^r a homeomorphic image S_0^n of S^n and a homeomorphic image S_1^{r-n-1} of S^{r-n-1}, where $r > n \geqq 0$ and $S_0^n S_1^{r-n-1} = 0$. Assigning orientations to S_0^n and S_1^{r-n-1} we obtain then a linkage coefficient whose absolute value, denoted by $\bar{v}(S_0^n, S_1^{r-n-1})$, is independent of the chosen orientations. We have proved in another paper[11] that S_0^n is a retract of $S^r - S_1^{r-n-1}$ if and only if $\bar{v}(S_0^n, S_1^{r-n-1}) = 1$. We shall now discuss the analogous problem obtained by replacing retracts by deformation retracts.

THEOREM 9. *Let S_0^n and S_1^{r-n-1} be disjoint homeomorphic images of S^n and S^{r-n-1} contained in S^r, where $r > n > 0$.*

If $n > 1$ then S_0^n is a deformation retract of $S^r - S_1^{r-n-1}$ if and only if $\bar{v}(S_0^n, S_1^{r-n-1}) = 1$ and $\pi_1(S^r - S_1^{r-n-1}) = 0$.

If $n = 1$ then S_0^1 is a deformation retract of $S^r - S_1^{r-2}$ if and only if $\bar{v}(S_0^1, S_1^{r-2}) = 1$, $\pi_1(S^r - S_1^{r-2})$ is cyclic infinite, and if $S^r - S_1^{r-2}$ is i-simple for $i = 2, 3, \cdots, r$.

PROOF. If S_0^n is a deformation retract of $S^r - S_1^{r-n-1}$ then by the theorem quoted above $\bar{v}(S_0^n, S_1^{r-n-1}) = 1$. It also follows that S_0^n and $S^r - S_1^{r-n-1}$ have the same homotopy type and therefore the remaining conditions of Th. 9 are fulfilled in view of Th. 8.

Now, let $\bar{v}(S_0^n, S_1^{r-n-1}) = 1$ and let r be a retraction mapping $S^r - S_1^{r-n-1}$ into S_0^n. Let γ_0^n be a basis-cycle for the group ${}^{n}\mathcal{H}^{I}(S_0^n)$. We shall prove that γ_0^n is also a basis-cycle for ${}^{n}\mathcal{H}^{I}(S^r - S_1^{r-n-1})$. In fact, if this is not the case there is a singular cycle γ^n in $S^r - S_1^{r-n-1}$ such that $\gamma_0^n \sim k\gamma^n$ in $S^r - S_1^{r-n-1}$

[11] *Fund. Math.* 31 (1938), p. 192, see also K. Borsuk and S. Eilenberg, *Fund. Math.* 26 (1936), p. 215.

where $k \neq \pm 1$. This implies $\gamma_0^n = r(\gamma_0^n) \sim kr(\gamma^n)$ in S_0^n contradicting our hypothesis that γ_0^n was a basis-cycle for ${}^n\mathcal{H}^I(S_0^n)$. Therefore γ_0^n is a basis-cycle for ${}^n\mathcal{H}^I(S^r - S_1^{r-n-1})$. Since $r(\gamma_0^n) = \gamma_0^n$ it follows that $h_r^n = h_1^n$, where 1 is the identity mapping of $S^r - S_1^{r-n-1}$ into itself.

If the remaining conditions of Th. 9 are also fulfilled then, by Th. 8, $S^r - S_0^{r-n-1}$ has the homotopy type of S_0^n. From Th. 5 (or Th. 5_1) it then follows that (2.2) of Th. 2 can be applied to mappings of $S^r - S_1^{r-n-1}$ into itself. Since $h_r^n = h_1^n$, we have $r \simeq 1$ so that S_0^n is a deformation retract of $S^r - S_0^{r-n-1}$.

10. If we take $n = 0$ in Th. 8 and Th. 9 then $S^r - S_1^{r-1}$ is not connected but consists of two connected regions C_1 and C_2. By duality theorems we have ${}^i\mathcal{H}^I(C_j) = 0$ for $i = 1, 2, \cdots$ and $j = 1, 2$. If we further admit that $\pi_1(C_j) = 0$ then by a theorem of Hurewicz[4] C_j can be deformed to a point. It follows that $S^r - S_1^{r-1}$ has then the homotopy type of S^0 where S^0 consists of two points. It is also clear that then every $S_0^0 \subset S^r - S_1^{r-1}$ such that $\bar{v}(S_0^0, S_1^{r-1}) = 1$ is a deformation retract of $S^r - S_1^{r-1}$. Hence we see that

(10.1) *The statements of Th.* 8 *and Th.* 9 *concerning the case* $n > 1$ *hold also for* $n = 0$ *provided the relation* $\pi_1(S^r - S_1^{r-1}) = 0$ *is interpreted as* $\pi_1(C_1) = \pi_1(C_2) = 0$ *where* C_1 *and* C_2 *are the components of* $S^r - S_1^{r-1}$.

In the case $n = 1$ Theorems 8 and 9 contain the condition that $S^r - S_1^{r-2}$ should be i-simple for $i = 2, 3, \cdots, r$. In the case $r = 3$ (and also in the trivial case $r = 2$) the author has proved[12] that this condition can be dropped. The similar question concerning $r > 3$ remains unanswered and seems to be closely related to the following problem:

Let S_1^{r-n-1} be a homeomorphic image of S^{r-n-1} in S^r where $r > n > 0$. Under what conditions is $S^r - S_1^{r-n-1}$ aspherical (i.e. $\pi_i(S^r - S_1^{r-n-1}) = 0$ for $i > 1$)?

In particular, is $S^r - S_1^{r-n-1}$ aspherical if $n = 1$ and $\pi_1(S^r - S_1^{r-2})$ is cyclic infinite?

Again this question has been answered positively only for $r = 2, 3$.[12]

UNIVERSITY OF MICHIGAN

[12] *Fund. Math.* 28 (1937), p. 238.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

ON PROJECTIVE EQUIVALENCE OF TRILINEAR FORMS

BY ROBERT M. THRALL

(Received March 20, 1940)

1. Introduction. The two fundamental problems in the projective classification of trilinear forms are: (I) the determination of necessary and sufficient conditions for the (projective) equivalence[1] of two trilinear forms; (II) the determination of a set of forms which consists of one and just one representative form from each *class* of equivalent forms.

In the form[2] $F(x, y, z) = a_{\alpha\beta\delta}y_\alpha z_\beta x_\delta$ let there be py's, qz's, and rx's and suppose that the numbers p, q, r are the two way rank invariants[3] of the form F. Let the variables be so named that $p \leqq q \leqq r$. We can measure progress in the solution of problems I and II in terms of the sets of invariants (p, q, r) for which they have been solved.

Problems I and II have been rationally solved (i.e. for the coefficients $a_{\alpha\beta\delta}$ belonging to any field) for $p = 1$ and $p = 2$. For if $p = 1$, then $q = r$ and is the only invariant of F. For $p = 2$ the trilinear form can be considered as a pencil of bilinear forms. Two such pencils are equivalent if and only if they have the same minimal numbers and projectively equivalent invariant factors.[4]

If $p > 2$ problems of rationality enter, and most of the progress here has been for forms with coefficients in the field, K, of complex numbers; although in [13], the case $(p, q, r) = (3, 3, 3)$ was discussed for forms belonging to a Galois field by rational methods applicable to any field. [14] contains the complete solution for forms (3, 3, 3) with coefficients in K.

In §§2–9 of the present paper, methods similar to those used in [14] are applied to the next case (3, 3, 4). The coefficient field is K although many of the results are obviously more general. §2 is devoted to definitions and notation for determinantal manifolds associated with trilinear forms. In §3 the forms (3, 3, 4) are divided into generate and non-degenerate cases according to their associated determinantal manifolds. §§4–8 give classification theorems for the various non-degenerate cases. Theorem 4.3 is basic in this discussion. §5 discusses briefly generalization of 4.3 to forms $(p, p, 2p - 2)$. The principal results of these sections are found in theorems 6.5, 6.6, 7.1, 8.6. §9 (especially theorem 9.1) and the appendix contain the solutions of problems I and II for the degen-

[1] For definitions of equivalence and g-equivalence, see [14] p. 678. (The numbers in brackets refer to the bibliography.)

[2] Repeated Greek indices indicate summation.

[3] This means that F is equivalent to no form with less than py's, qz's and rx's. For discussion of and reference to these invariants, see [14] p. 678.

[4] See [9], and [12] p. 682.

erate cases. In §10 the work of T. G. Room[5] on determinantal manifolds is applied to the "general" case for $3 \leqq p < q \leqq r$ and $3 < p = q \leqq r$ to obtain theorems which reduce problem I to a study of projective equivalence of manifolds related to the forms. Also in the section is a discussion of different interpretations of the concept "general" as applied to trilinear forms.

2. The determinantal manifolds[6] associated with a trilinear form. Let $F(x, y, z) = a_{\alpha\beta\delta}y_\alpha z_\beta x_\delta \equiv y_\alpha z_\beta x_{\alpha\beta} \equiv z_\beta x_\delta y_{\beta\delta} \equiv y_\alpha x_\delta z_{\alpha\delta}$ define the elements of the matrices $\| x_{\alpha\beta} \|$, $\| y_{\beta\delta} \|$, $\| z_{\alpha\delta} \|$. Denote by $V_x^s = (| p, q |_s[r-1])$ the manifold consisting of points x in projective $r-1$ space for which $\| x_{\alpha\beta} \|$ is of rank s or less. We write the equations for V_x^s in the form $| x_{\alpha\beta} |_s = 0$. We call $V_x = V_x^{p-1}$ *the principal manifold* of the matrix $\| x_{\alpha\beta} \|$ and of the form F.[7] The manifolds V_y^s and V_z^s are defined analogously from the matrices $\| y_{\beta\delta} \|$ and $\| z_{\alpha\delta} \|$. The equivalence of two trilinear forms obviously implies the projective equivalence of their corresponding manifolds[8] V, and in general the converse is also true. (See §10.)

Since the derivative of an n-rowed determinant is a linear combination of its $n-1$ rowed minors, the manifold V^{s-1} must be multiple on V^s. In general V_x^s has dimension $r - 1 - (q - s)(p - s)$ and so will necessarily be non vacuous only for values of s for which this dimension is positive. In particular cases, however, the dimension may be greater than this number.

3. Trilinear forms (3, 3, 4), **general theory.** In general, $V_x = V_x^2$ (the principal manifold), a cubic surface in 3-space, and V_y^2 and V_z^2, zero dimensional manifolds of order six in 2-space, are the only non empty manifolds associated with a trilinear form (3, 3, 4). However, there are degenerate cases in which some of the V^1's exist, or in which V_y^2 or V_z^2 is of dimension greater than zero. On this basis we divide our treatment into two parts:

(A) *The non-degenerate cases*: V_y^1 is empty, and V_y^2 is of dimension zero and order six.

(B) *The degenerate cases*: At least one of the conditions under (A) is not satisfied.

Before treating (A) and (B) separately we study three algebraic transformations defined by the form F. We say that $\bar{y}$ and $\bar{z}$ are images under the transformation T_1 if

$$a_{\alpha\beta\delta}\bar{y}_\alpha\bar{z}_\beta = 0, \qquad \delta = 1, 2, 3, 4. \tag{3.1}$$

The transformations T_2 and T_3 are similarly defined by

$$a_{\alpha\beta\delta}\bar{z}_\beta\bar{x}_\delta = 0, \quad \alpha = 1, 2, 3, \qquad \text{and} \qquad a_{\alpha\beta\delta}\bar{y}_\alpha\bar{x}_\delta = 0, \quad \beta = 1, 2, 3.$$

[5] [3] especially Chapter VII, or [10]. Chapter VII is based upon [10]. For a review of [3] see [11].

[6] The definitions and notation are those given by Room [3] Chapter I.

[7] If $r = q$ ($r = q = p$) F has two (three) principal manifolds, but if $r > q \geqq p$ there is just one principal manifold.

[8] See [13] p. 386.

Considered as linear equations in $\bar{z}$ the number of solutions of (3.1) depends upon the rank of $\| y_{\beta\delta} \|$, and inversely for $\bar{y}$ and $\| z_{\alpha\delta} \|$. Thus if $\bar{y}$ and $\bar{z}$ are images under T_1, then $\bar{y} \in V_y^2$ and $\bar{z} \in V_z^2$. If $\bar{y} \in V_y^1$ it will have a whole line of images $\bar{z}$, which must all lie on V_z^2. If $\bar{y} \in V_y^2$ but $\notin V_y^1$, then it will have a single image $\bar{z}$. Hence, if V_y^2 is of dimension zero, V_z^1 is empty; and, if further V_y^1 is empty, the points of V_y^2 and V_z^2 will be in 1-1 correspondence.

From this it follows that the conditions under (A) above are actually symmetric in y and z although stated in terms of y only. A zero dimensional manifold, V_y^2, of order six is, in general, a set of six distinct points. However, these points may coincide in various ways so that the actual number of distinct points in V_y^2 may be anything from one to six. The above arguments show that *under* (A) *if V_y^2 has k distinct points then V_z^2 likewise has k distinct points.* We shall show later that $\bar{y}$ has the same multiplicity in V_y^2 as its image $\bar{z}$ has in V_z^2.

T_2 is a transformation between points of V_z^3 (the whole z-plane) and points of V_x^2. If $\bar{z} \in V_z^1$ its image under T_2 is a plane, which must be a linear factor of V_x^2; if $\bar{z} \in V_z^2$ but $\notin V_z^1$ its image is a line on V_x^2; and if $\bar{z} \notin V_z^2$, its image is a single point of V_x^2. T_3 is of the same general character as T_2.

4. The non-degenerate cases. Consider the general trilinear form (3, 3, 4) in which the coefficients $a_{\alpha\beta\delta}$ are independent indeterminants. The three rowed minors of $\| y_{\beta\delta} \|$ define a web, W_y, of cubics whose base is the six (in this case distinct) points of V_y^2. Let $f_\delta(y)$ denote the three rowed minor of $\| y_{\beta\delta} \|$ obtained by dropping the δ-th column. Then two generic members $G(\lambda, y) = \lambda_\delta f_\delta(y)$ and $G(\mu, y)$ of W_y will meet in nine points of which six, say $y^{(1)}, \cdots, y^{(6)}$ have coordinates free of λ and μ. The u-resultant[9] of $G(\lambda, y)$ and $G(\mu, y)$ will therefore split rationally into a factor

$$H(u, a_{\alpha\beta\delta}) = \prod_{\gamma=1}^{6} y_\alpha^{(\gamma)} u_\alpha$$

of degree 6 and free of λ and μ, and a second factor of degree 3.

If now we give the $a_{\alpha\beta\delta}$ arbitrary values in the ground field K, for which $H(u, a_{\alpha\beta\delta}) \not\equiv 0$, some or all of the six points, $y^{(\gamma)}$, may coincide. If a point, P, of V_y^2 arises from an s-fold factor of H then we assign to P the multiplicity s in V_y^2, thus ensuring that V_y^2 will have multiplicity 6 in any form for which $H \not\equiv 0$. On the other hand this multiplicity s assigned to P is by its definition the same as the multiplicity of intersection at P of two generic cubics of the web W_y.

We now divide the non-degenerate cases into two sets:

The linear non-generate cases: W_y contains a curve of genus one.

The nodal non-degenerate cases: W_y contains no curve of genus one. In this section we shall treat only the linear cases.

If C is a curve of genus one belonging to W_y, we can utilize its Puiseux ex-

[9] See [5] vol. 2 p. 21.

pansions to characterize V_y^2. To obtain uniqueness we choose the particular expansion

$$y_\alpha = E_\alpha(t) \quad \alpha = 1, 2, 3$$

at a point $P = (c_1, c_2, c_3)$ for which

$$E_i(t) = 1, \quad E_j(t) = c_j/c_i + t, \quad E_k(t) = c_{k\gamma}t^\gamma \quad \gamma = 0, \cdots, \infty \qquad (c_{k0} = c_k/c_i)$$

where i is the smallest index for which $c_i \neq 0$, and $j < k$ unless a vertex of the coordinate triangle lies on the tangent to C at P, in which case $c_{k1} = 0$ and it may be necessary to take $j > k$. That this is possible follows from the linearity of every branch of C and the general theory of Puiseux expansions.[10] We shall call these specialized expansions the *canonical expansion* of C at P.

Let the web, W_y, be given by

$$G(\lambda, y) = \lambda_\delta f_\delta(y) \quad \delta = 1, \cdots, 4$$

If $G(\lambda, E(t)) \equiv 0 \bmod t^s$ but $\not\equiv 0 \bmod t^{s+1}$ then P counts s times in V_y^2. Let $P_\alpha(t)$ be obtained from $E_\alpha(t)$ by dropping all terms whose degree in t is s or greater. If C' is any other member of W_y (of genus one), then the first s terms of its canonical expansion $E'_\alpha(t)$ will likewise be $P_\alpha(t)$ and so the polynomials $P_\alpha(t)$ express properties of the web and not merely of the particular curve C in the web. We call the $P_\alpha(t)$ the *canonical polynomials* of W_y at P. (We note that these polynomials are identically zero unless $P \in V_y^2$.)

Let P^μ, $\mu = 1, \cdots, h$, count s_μ times in V_y^2, $(s_1 + \cdots + s_h = 6)$, and let the canonical polynomials of W_y at P^μ be $P^\mu_\alpha(t)$.

4.1 Definition. The curve $K: g(y) = 0$ is said to pass through V_y^2 if and only if $g(P^\mu(t)) \equiv 0 \bmod t^{s_\mu}$ $\mu = 1, \cdots, h$.

4.2 Theorem. *The cubic $g(y) = 0$ belongs to W_y if and only if it passes through V_y^2.*

For there is just one web of cubics satisfying the six conditions of definition 4.1. (4.2 does not hold in the nodal cases.)

The results of this section obviously hold for the z-space as well as the y-space. The following theorem is fundamental in the solution of problem II for the linear case and with a slight modification applies also to linear branches of V_y^2 in the non linear case.

4.3 Theorem. *Let F be a linear non-degenerate form and let T_1 send P on V_y^2 into P' on V_z^2. If a branch of V_y^2 at P has multiplicity σ, with canonical polynomials $P_\nu(t)$, then there exists a linear branch of V at P' of multiplicity σ with canonical polynomials $P'_\nu(\tau)$, and a polynomial $\tau = \tau(t) = b_1t + b_2t^2 + \cdots + b_{\sigma-1}t^{\sigma-1}$, $b_1 \neq 0$, such that*

$$(B_\sigma)\colon\ F(t, x_{\alpha\beta}) = P_\alpha(t)P'_\beta(\tau(t))x_{\alpha\beta} \equiv 0 \bmod t^\sigma.$$

[10] See for instance [2] pp. 213–231.

PROOF. We take $P = (c_1, c_2, c_3)$, $P' = (c_1', c_2', c_3')$ and for convenience suppose $c_1c_1' \neq 0$.

The four congruences, B_σ, can be regarded as homongeneous linear equations with coefficients in the ring, R, of polynomials in t reduced modulo t^σ. The matrix $\| y_{\beta\delta}P_\alpha(t) \|$ is by hypothesis of rank 2 (in R). Hence there exists a solution $z_\beta = Q_\beta'(t)$ with the following properties:[11] $z_1 = 1$, $z_2 = c_2'/c_1' + E_2(t)$, $z_3 = c_3'/c_1' + E_3(t)$ where the $E_\beta(t)$ belong to R (i.e. are polynomials in t of degree less than σ).

The polynomials $P_\alpha(t)$, not all zero divisors ($P_1(t) = 1$), are solutions of the congruences $z_{\alpha\delta}(Q_\beta'(t))y_\alpha = 0$. Hence the rank of $\| z_{\alpha\delta}(Q_\beta'(t) \|$ is <3 (actually 2 since the case is non-degenerate). To complete the proof of the theorem we need merely to show the existence of a $\tau = b_1t \cdots + b_{\sigma-1}t^{\sigma-1}$, $b_1 \neq 0$ such that $Q_\beta'(t) = P_\beta'(\tau(t))$ where the $P_\beta'(\tau)$ are "canonical" polynomials for V_z^2. If $E_2(t) \not\equiv 0 \bmod t$ we take $\tau = E_2(t)$. Otherwise take $\tau = E_3(t)$. [If both $E_2(t)$ and $E_3(t)$ were divisible by t^2 the whole line $y_k - c_{k1}y_j = (c_k - c_jc_{k1})y_1$ would map into P' under T_1 contrary to the hypothesis of non-degeneracy.]

5. Linear non-degenerate forms $(p, p, 2p - 2)$. The forms (3, 3, 4) belong to the series of forms $(p, p, 2p - 2)$ in which the principal manifold is a determinant, and in which V_y^{p-1} and V_z^{p-1} are, in general, zero dimensional, of order $N = \binom{2p-2}{p-1}$. A form $(p, p, 2p - 2)$ is said to be *non-degenerate* if V_y^{p-1} is zero dimensional of order N, and V_y^{p-2} is empty. A non-degenerate form $(p, p, 2p - 2)$ is said to be *linear* if V_y^{p-1} can be defined by canonical polynomials satisfying 4.1 with the word "curve" replaced by "hypersurface." With these definitions theorem 4.3 is valid for forms $(p, p, 2p - 2)$ if we replace V_y^2, V_z^2 by V_y^{p-1}, V_z^{p-1} throughout. The proof of this generalization is stepwise parallel to that of 4.3.

6. The general non-degenerate case. In the general case V_y^2 consists of six distinct points $y^{(1)}, \cdots, y^{(6)}$, not on a conic and no three on a line. This implies the same for V_z^2, since otherwise T_3 and T_2 could not map the y and z planes respectively into the same surface V_x^2. Let coordinates be so chosen that $y^{(1)} = (1, 0, 0)$, $y^{(2)} = (0, 1, 0)$, $y^{(3)} = (0, 0, 1)$, $y^{(4)} = (1, 1, 1)$, $y^{(5)} = (e, a, b)$, $y^{(6)} = (e, c, d)$; $z^{(1)} = (1, 0, 0)$, $z^{(2)} = (0, 1, 0)$, $z^{(3)} = (0, 0, 1)$, $z^{(4)} = (1, 1, 1)$, $z^{(5)} = (e, a', b')$, $z^{(6)} = (e, c', d')$; and so that

$$(6.1) \qquad y_\alpha^{(i)} z_\beta^{(i)} x_{\alpha\beta} \equiv 0 \qquad i = 1, \cdots, 6$$

expresses the congruence of V_y^2 and V_z^2 under T_1. Of these six equations, one, say the last, must be dependent on the other five; for otherwise there would be

[11] The existence theorem for linear equations in a commutative ring assures the existence of a solution with not all Q_β' null divisors. By taking $\sigma = 1$ we see that Q_1' is not a null divisor (since $c_1' \neq 0$). But every non-null divisor in R has an inverse in R, so that $z_\beta = Q_\beta'/Q_1'$ is also a solution and has $z_1 = 1$.

only three independent $x_{\alpha\beta}$, whereas $r = 4$. This means that the coordinates of $y^{(6)}$ and $z^{(6)}$ are rational functions of the coordinates of $y^{(i)}$, $z^{(i)}$, $i < 6$.

$a \neq b$, for if $a = b$ then $y^{(1)}$, $y^{(4)}$, $y^{(5)}$ are collinear. Hence we can solve the first five equations of 6.1 for x_{21} and x_{31} giving

$$x_{11} = x_{22} = x_{33} = 0, \quad x_{21} = \alpha_1 x_{12} + \alpha_2 x_{13} + \alpha_3 x_{23} + \alpha_4 x_{32},$$

$$x_{31} = (-1 - \alpha_1)x_{12} + (-1 - \alpha_2)x_{13} + (-1 - \alpha_3)x_{23} + (-1 - \alpha_4)x_{32}$$

where

$$\alpha_1 = \frac{b - a'}{a - b}, \qquad \alpha_2 = \frac{b - b'}{a - b}, \qquad \alpha_3 = \frac{eb - ab'}{e(a - b)}, \qquad \alpha_4 = \frac{(e - a')b}{e(a - b)}.$$

If now we set $x_{12} = x_1$, $x_{13} = x_2$, $x_{23} = x_3$, $x_{32} = x_4$ and compute V_y^2 we find

$$(6.2) \qquad y^{(6)} = [e(a - a')(b - b')(a' - b'), (a - a')(b' - e)(ab' - a'b),$$
$$(b - b')(a' - e)(ab' - a'b)],$$

and from the symmetry we obtain $z^{(6)}$ by substituting z, a', b', a, b for y, a, b, a', b' in 6.2.

6.3 THEOREM. *Given $y^{(i)}$, $i < 6$, and $z^{(i)}$, $i < 5$ there exists a unique point $z^{(5)}$ for which $y^{(6)}$ is any preassigned point (e, c, d) not on the conic or any of the lines determined by $y^{(i)}$, $i < 6$.*

For proof it is sufficient to observe that the equations

$$(6.4) \qquad \begin{aligned} \rho t_1' &= (at_1 - et_2)(bt_1 - et_3)(t_2 - t_3) \\ \rho t_2' &= (at_1 - et_2)(t_3 - t_1)(at_3 - bt_2) \\ \rho t_3' &= (bt_1 - et_3)(t_2 - t_3)(at_3 - bt_2) \end{aligned}$$

define an involutorial plane Cremona transformation. If now we set $t' = y^{(6)}$ and $t = z^{(5)}$, 6.4 reduces to 6.2, so that 6.2 has a unique solution for a', b' in terms of e, a, b, c, d. (The excluded conic and lines include all of the principal curves of the transformation.)

Theorem 6.3 enables us to give the complete solution of problems I and II in the general case. For we have at once the following theorems:

6.5 THEOREM. *There exists a trilinear form with $(p, q, r) = (3, 3, 4)$ for which V_y^2 is any preassigned, general set of six points in the y plane.*

6.6 THEOREM. *Two forms, F and $\bar{F}$, for which V_y^2 is general are equivalent if and only if V_y^2 is projectively equivalent to $\bar{V}_y^2$.*

The projective invariants of the six point plane manifold have been given by Coble in [7].

6.7 COROLLARY. *Two non-equivalent general forms F and $\bar{F}$ are g-equivalent if and only if V_y^2 is projectively equivalent to $\bar{V}_z^2$.*

Hence, the g-class containing a form is identical with the class containing the form if and only if V_y^2 is projectively equivalent to V_z^2 (not necessarily in the

order of congruence under T_1). In general this is not the case, but if $e, a, b, c, d,$ satisfy certain syzygies it may happen.

7. The non-general, non-degenerate linear cases. There are approximately a hundred different projective types of special zero dimensional manifolds of order 6. For instance, there may be six distinct points; all on a conic, three on a line, three on one line and the other three on a second line, five on two lines meeting at one of the points, the vertices of a triangle plus a point on each side, the vertices of a quadrilateral.[12] If two of the points coincide (become "infinitely near") along some fixed direction, there are 11 types of ways in which the other four points can be placed relative to the first point and the given direction at it, etc. etc. We shall not consider each such case separately, but the following theorem enables one to determine when two such manifolds U and U' can serve as V_y^2 and V_z^2 for a trilinear form, thus giving the information needed to reduce problem II (for the forms of this section) to a matter of direct computation.

7.1 THEOREM. *If U and U' are two zero dimensional (non-nodal) manifolds of order* 6 *there exists a trilinear form* (3, 3, 4) *with $V_y^2 = U$ and $V_z^2 = U'$ if and only if the following conditions are satisfied.*

a) *the points and branches of U and U' can be put into* 1-1 *correspondence in accordance with the requirements of theorem* 4.3.

b) *the resulting relations B_σ considered as linear equations in the $x_{\alpha\beta}$ shall have a matrix of rank* 5.

c) *the V_x^2 thus defined is irreducible, and without a line of nodes.*

PROOF: Necessity, a) is a consequence of theorem 4.3; b), there are just 9 $x_{\alpha\beta}$ of which just 4 $(=r)$ are independent; c), if V_x^2 were reducible it could not be the rational map of the y plane and so one of its factors would have to be the image of a one dimensional part of V_y^2 or of a point of V_y^1, either possibility giving a degenerate case. If V_x^2 has a line of nodes its plane sections are rational. Hence, the corresponding curves in W_y must be rational; and the case is either degenerate or nodal.

Sufficiency. a) and b) are sufficient to define a form (3, 3, 4) with V_y^2 including U and V_z^2 including U', with the inclusion being equality unless the case is degenerate. Reference to §9 shows that V_x^2 contains a line of nodes in the degenerate cases where it is irreducible. Hence *c*) implies that the case is non-degenerate (as well as non-nodal).

This theorem is valid for the general case but is weaker than theorems 6.3 and 6.5. The following extension of 6.6, which solves problem I for all non-degenerate, non-nodal cases, is an immediate consequence of 7.1.

7.2 THEOREM. *Two non-degenerate, non-nodal forms F, $\bar{F}$ are equivalent if and only if V_y^2 and V_z^2 are projectively equivalent to $\bar{V}_y^2$ and $\bar{V}_z^2$, respectively.*

[12] If four of the points are on a line, the case is degenerate with the whole line belonging to V_y^2.

The following two lemmas are useful in applications of 7.1.

7.3 LEMMA. *If a line λ cuts V_y^2 3 times in points whose images under T_1 lie on a line λ', then either the case is nodal; or V_y^2 includes λ, V_z^2 includes λ', and the V_x^2 is reducible.* (It is not necessary that λ cut V_y^2 in 3 distinct points.)

7.4 LEMMA. *If V_y^2 lies on a conic, then the case is degenerate.*

Both of the proofs follow upon direct application of 4.3 and will be omitted.

7.1 and 7.3 enable us to prove the following theorem:

7.5 THEOREM. *Not every quaternary cubic can be written as a three rowed determinant whose elements are linear forms.*[13]

More specifically and in the language of this paper the theorem is

7.6 THEOREM. *V_x^2 can be any quaternary cubic excepting one projectively equivalent to $f(x) = x_1^3 + x_2^2x_3 + x_3^2x_4$, characterized projectively by its unode of the third kind.*[14]

PROOF. The surface $f(x) = 0$ contains just one line, μ. Suppose that V_y^2 contains the two points P_1 and P_2 both of which map into μ under T_3. Then each point of μ has two images in the y plane and therefore all of μ lies on V_x^1, i.e. μ is a line of nodes of V_x^2. But $f(x)$ has only one node. We have proved that if V_x^2 is $f(x) = 0$ then V_y^2 has just one point, P. P cannot be a node of W_y for then every plane section of V_x^2 would be rational.

Next suppose that P is not a point of inflexion for all members of the web, W_y. Then both P and the tangent at P to the curves of the web must map into μ under T_3, which would again require μ to be a line of nodes.

The only remaining possibility is that V_y^2 be a single point at which the curves of the mapping web have a common inflexion tangent, λ, and three further "infinitely near" intersections. Similarly, V_z^2 is a single point cut three times by a line λ'. But now λ and λ' satisfy the hypotheses of 7.3, and so by 7.1 no determinantal representation of $f(x)$ can exist. To complete the proof we remark that representations for the other projective types can be readily constructed.

We conclude this section with some applications of 7.1 which illustrate the general method of attack in the special, linear, non-degenerate cases. We ask if W_y can be a web of cubics with a common inflexion tangent at one point and with a triple intersection at a second point; and the same for W_z.

Choose y coordinates so that W_y has the inflexion point (1, 0, 0) with tangent $y_3 = 0$, and so that the second base point is (0, 0, 1) with common tangent $y_1 = 0$. Then the canonical polynomials will be $P^{(1)} = (1, t, 0)$ and $P^{(2)} = (at^2, t, 1)$. Then choose z coordinates so that the canonical polynomials are $Q^{(1)} = (1, \tau_1, b\tau_1^2)$ and $Q^{(2)} = (0, \tau_2, 1)$.

If $Q^{(2)}$ and $P^{(1)}$ were images under T_1, then by 7.3 the case would be degenerate. Hence, $P^{(i)}$ corresponds to $Q^{(i)}$, $i = 1, 2$. We have two sets of equations B_σ,

[13] In [8] p. 175 Dickson has proved that every "sufficiently general" quaternary cubic can be written as a determinant, but he gives no counter example. Room ([3] p. 65) states, incorrectly, that any quaternary cubic is determinantal.

[14] See [1] p. 61.

each with $\sigma = 3$. If we can choose constants $b_1 \neq 0, b_2;\ c_1 \neq 0, c_2$ so that $\tau_1 = b_1 t + b_2 t^2$ and $\tau_2 = c_1 t + c_2 t^2$ we will have met condition a) of 7.1. The equations are:

$$B_3^{(1)} \qquad \begin{cases} x_{11} = 0 \\ x_{21} + b_1 x_{12} = 0 \\ b_1 x_{22} + b_2 x_{12} + b b_1^2 x_{13} = 0 \end{cases}$$

$$B_3^{(2)} \qquad \begin{cases} x_{33} = 0 \\ x_{23} + c_1 x_{32} = 0 \\ a x_{13} + c_1 x_{22} + c_2 x_{32} = 0. \end{cases}$$

Necessary and sufficient conditions for the dependence of these equations are $b_2 = c_2 = 0$, $c_1 b_1^2 b = a b_1$ or $c_1 = a/bb_1$. This gives

$$\| x_{\alpha\beta} \| = \begin{Vmatrix} 0 & x_{12} & x_{13} \\ -b_1 x_{12} & -bb_1 x_{13} & x_{23} \\ x_{31} & -(a/bb_1)x_{23} & 0 \end{Vmatrix} \sim \begin{Vmatrix} 0 & x_1 & x_2 \\ -x_1 & -x_2 & x_3 \\ x_4 & -x_3 & 0 \end{Vmatrix}.$$

Then $V_x^2 : x_1 x_3 (x_2 + x_4) + x_2^2 x_4 = 0$ is irreducible and has only three nodes so that condition c) of 7.1 is satisfied, and the existence of the form in question is proved. Incidentally, the canonical form obtained shows that there is but one class of forms having V_y^2 and V_z^2 of the above required type.

We can show, further, that for this particular choice of V_y^2 there can be only the one choice of V_z^2 (to within projectivities). For by 7.1 V_z^2 must have just two points, each of multiplicity three. By a purely geometric study of the surface V_x^2 defined by V_y^2, it can be shown that there are just two possible kinds of mapping webs having two base points each of multiplicity 3. One of these is V_y^2 and the other has a base lying on a conic and so is excluded by 7.4.

However, it is not always necessary that V_y^2 and V_z^2 be of the same projective type. For consider the form represented by

$$\| x_{\alpha\beta} \| = \begin{Vmatrix} 0 & x_{12} & x_{13} \\ x_{12} & x_{13} + a x_{12} & x_{31} \\ x_{31} & x_{32} & 0 \end{Vmatrix}.$$

V_y^2 has canonical polynomials $P^{(1)} = (1, t, -t^3)$, $P^{(2)} = (0, t, 1)$ and V_z^2 has $Q^{(1)} = (1, \tau_1, \tau_1^2 - a\tau_1^3)$, $Q^{(2)} = (\tau_2, 0, 1)$ where $\tau_1 = -t + at^2 - a^2 t^3$, $\tau_2 = -t$. The members of W_y have a common inflexion tangent whereas those of W_z do not. We note that for a form $\bar{F}$ to have both $\bar{V}_y^2$ and $\bar{V}_z^2$ projectively equivalent either to V_y^2 or to V_z^2 would contradict 7.3.

8. The nodal non-degenerate cases. A linear system of curves has no variable multiple point.[15] Hence, for every member of W_y to be of genus zero there must

[15] Bertini's theorem. [4] p. 25.

be a fixed double point (cusp). This double point counts four times as a member of V_y^2 but it is only three (linear) conditions on a cubic to have a fixed double point. For a cubic to contain the remaining two points (or branches) of V_y^2 is at most two further linear conditions. Hence, there is a $10 - 5 = 5$ parameter family of cubics on V_y^2 of which W_y is some 4 parameter subfamily.

We see readily that there are just five projective types of nodal cases:

N_1 : node plus two outside points.

N_2 : node with one fixed tangent plus an outside point.

N_3 : node with both tangents fixed.

N_4 : node with one tangent and a further fixed direction along that branch.

N_5 : cusp (the tangent must then be fixed).

Let t and u be indeterminants over K and set $R = K(u)[t] \bmod t^2$.

8.1 THEOREM. *If P and P' are images under T_1, then P is a node of W_y if and only if there exist polynomials linear in t and in u such that*

$$F(t, x_{\alpha\beta})P_\alpha P'_\beta x_{\alpha\beta} \equiv 0 \bmod t^2 \text{ identically in } u; \tag{8.2}$$

furthermore, if P is a node in W_y, then P' is a node in W_z.

PROOF: We suppose y and z coordinates so chosen that $P = (1, 0, 0)$ and $P' = (1, 0, 0)$. If P is a node of W_y, then any line $\rho y_1 = 1 = P_1$, $\rho y_2 = t = P_2$, $\rho y_3 = ut = P_3$, through P must meet the curves of W_y twice at P. Or stated analytically, if W_y is $G(\lambda, y) = \lambda_\delta f_\delta(y)$ then $G(\lambda, P_\alpha) \equiv 0 \bmod t^2$ identically in u.

Now since $\| y_{\beta\delta}(P_\alpha) \|$ is of rank 2 (in R), we proceed as in the proof of theorem 4.3 to obtain solutions $z_\beta = P'_\beta(t)$ of the equations 8.2 such that $P'_1 = 1$, $P'_2 = tf_2(u)$, $P'_3 = tf_3(u)$ where the f_β are rational functions of u with coefficients in K. A short argument (which we omit) shows that the form is degenerate unless f_2 and f_3 are linearly independent first degree polynomials in u, say $f_\beta(u) = c_\beta u + d_\beta$, $\beta = 1, 2$. The polynomials P, P' satisfy the conditions of the theorem. Further, since $\rho z_\beta \doteq P'_\beta$ is by suitable choice of u, any line through P', 8.2 is symmetric in y and z so that P' is a node of W. This completes the proof of the necessity of 8.2. The sufficiency is obvious.

The cases $N_1, \cdots, N_4$ are sufficiently alike to justify our treating only one, N_1, in detail and merely listing a representative of each class in the other cases. Suppose then that V_y^2 consists of the points (1, 0, 0), (0, 1, 0), (0, 0, 1) of which the first is a node and let T_1 map these points into $z = (1, 0, 0), (0, 1, 0), (0, 0, 1)$ respectively. Now apply 8.2 and 4.3. We obtain

$$\| x_{\alpha\beta} \| = \left\| \begin{matrix} 0 & x_{12} & x_{13} \\ d_2 x_{12} + d_3 x_{13} & 0 & x_{23} \\ c_2 x_{12} + c_3 x_{13} & x_{32} & 0 \end{matrix} \right\|. \tag{8.4}$$

Let $x_{12} = x_2$, $x_{13} = x_3$, $x_{23} = x_1$, $x_{32} = x_4$. Then V_x^2 is

$$x_1 x_2(c_2 x_2 + c_3 x_3) + x_3 x_4(d_2 x_2 + d_3 x_3) = 0.$$

V_x^2 contains a line, l_n, of nodes, with the equations $x_2 = x_3 = 0$. The members of the pencil of planes $x_2 = \alpha x_3$ cut V_x^2 twice in l_n and in a further line

$$l_\alpha : x_1\alpha(\alpha c_2 + c_3) + x_4(\alpha d_2 + d_3) = 0, \quad x_2 = \alpha x_3 .$$

l_n and l_α meet at $(-\alpha d_2 - d_3, 0, 0, \alpha^2 c_2 + \alpha c_3)$. The point $(x_1, 0, 0, x_4)$ is on two distinct lines, l_α, l_{α^1}, unless the roots α, α^1 of

$$x_1(\alpha^2 c_2 + \alpha c_3) + x_4(\alpha d_2 + d_3) = 0$$

coincide. This happens if

$$(8.5) \qquad (c_3x_1 + d_2x_4)^2 - 4x_1x_4 d_3c_2 = 0.$$

The discriminant of 8.5 is

$$16c_2d_3 \begin{vmatrix} c_2 & c_3 \\ d_2 & d_3 \end{vmatrix}$$

which is zero only if the case is degenerate. Hence, 8.5 defines two distinct points, Q_1 and Q_2, on l_n. The points $R_1 = (1, 0, 0, 0)$ and $R_2 = (0, 0, 0, 1)$ are the only points of V_x^1.

8.6 THEOREM. *Two forms F and $\bar{F}$ belonging to case N_1 are g-equivalent if and only if the unordered pair of points Q_1, Q_2 are projectively equivalent to the unordered pair $\bar{Q}_1$, $\bar{Q}_2$ under a projectivity which sends V_x^1 into $\bar{V}_x^1$.*

PROOF: The "only if" is obvious since the points Q and R are defined purely in terms of V_x^2 and V_x^1. The "if" follows by normalization of 8.4 to a form containing a single parameter and then computation of the cross ratio of the four points in terms of this parameter. Subcases may arise in which one or both of the Q's may lie on V_x^1.

N_2 gives the single class represented by

$$||x_{\alpha\beta}|| = \left\| \begin{matrix} 0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \\ x_1 + x_2 & x_4 & 0 \end{matrix} \right\|.$$

N_3 gives the single class represented by

$$||x_{\alpha\beta}|| = \left\| \begin{matrix} 0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \\ x_2 & x_4 & x_1 \end{matrix} \right\|.$$

N_4 gives a family of g-classes with parameter $a \neq 0, -1$

$$||x_{\alpha\beta}|| = \left\| \begin{matrix} 0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \\ x_2 & ax_3 & x_4 \end{matrix} \right\|.$$

Two forms F, F' in case N_4 are g-equivalent if and only if $a + 1/a = a' + 1/a'$.

The argument for the cusp case is only slightly different. Necessary and sufficient conditions for W_y to be cuspidal can be written in the form of 4.3 and

8.2 utilizing Puiseux expansions. There is just one class under N_5, represented by

$$||x_{\alpha\beta}|| = \left\| \begin{matrix} 0 & x_1 & x_2 \\ -x_1 & x_2 & x_3 \\ -x_2 & -x_3 & x_4 \end{matrix} \right\|.$$

In every nodal case the principal manifold contains a line of nodes, and it is also a cone in cases N_3 and N_5.

9. The degenerate cases. Consideration of the manifolds V_y^1, V_z^1, V_x^1, V_y^2, V_z^2, V_x^2, and the relations between them defined by the transformations T_1, T_2, T_3 is sufficient to distinguish between the classes of degenerate forms in all save one set of classes, whose members are identical save for different values of a parameter and the distinctness of these classes follows from [14], p. 687; case 4.

The classes are first divided into sets of classes according to the nature of V_y^1 and V_z^1. Then among the classes possessing equivalent V_y^1 and equivalent V_z^1 we make a second subdivision according to V_x^1 or V_x^2 or both, and if this is still not sufficient to complete the classification we consider V_y^2 and V_z^2. In the appendix is a table which contains the matrix $|| y_{\beta\delta} ||$ for a representative of each degenerate class (3, 3, 4) and sufficient information to distinguish between the different classes. The proof of the completeness of the table is a normalization process along the lines of the similar proofs in [13] and [14] (i.e. successive discarding of forms equivalent to one retained). Because of its similarity to those and its length we shall omit the completeness proof.

We note that the terminology "degenerate cases" is justified in that V_x^2 is a reducible cubic in all of the degenerate classes except $g11$ and $g12$ in which it is a ruled cubic with a nodal line. Furthermore, it is also true that no manifolds can be V_x^2 for both a degenerate and a linear non-degenerate class, so that it would have been possible to have defined "degenerate class" in terms of V_x^2 alone, proper account being taken of classes $g11$ and $g12$.

As a partial summary of the results listed in the table of classes we give the following theorem which solves problem I for the degenerate cases.

9.1 THEOREM. *A necessary and sufficient condition for the equivalence of two degenerate trilinear forms F and $\bar{F}$ with $(p, q, r) = (3, 3, 4)$ is that V_t^1 and V_t^2 be simultaneously projectively equivalent to $\bar{V}_t^1$ and $\bar{V}_t^2$ for $t = x, y, z$ in turn, except in the following cases:*

a) *If V_x^1, V_y^1, V_z^1 consist of* 2, 1, 1 *points respectively, V_x^2 of three independent planes, and T_3, T_2 map V_y^1, V_z^1 into π_1, π_2 respectively,*[16] *and the same for $\bar{F}$, then $F \sim \bar{F}$ if and only if* i) $\pi_1 = \pi_2$ *and* $\bar{\pi}_1 = \bar{\pi}_2$ *or* ii) $\pi_1 \neq \pi_2$ *and* $\bar{\pi}_1 \neq \bar{\pi}_2$.

b) *If V_x^1, V_y^1, V_z^1 consist of* 1, 0, 0 *points respectively and V_y^2 is a line; and the same for $\bar{F}$, then $F \sim \bar{F}$ if and only if $a + 1/a = \bar{a} + 1/\bar{a}$ where a, $\bar{a}$ are the parameters of F and $\bar{F}$ in the canonical form given for the classes g_12 in the table.*

[16] π_1, π_2 being planes of V_x^2.

Of the 53 degenerate cases (the infinite family of g-classes discussed in 9.1b are counted as one case, g_12, but each other case includes only a single g-class) consideration of V_y^1, V_z^1, V_x^2, and V_x^1 is sufficient to isolate 37 and to divide the remaining 16 into 8 pairs. In four of these pairs: e_18, e_38; e_58, e_68; e_16, e_56; f_12, f_72 the distinctness is a consequence of different projective situations of V_x^1 on V_x^2. The distinction between e_36 and e_66 is given in 9.1a. In the remaining three pairs; f_32, f_42; f_21; f_31; g_22, g_32 consideration of V_y^2 and V_z^2 is sufficient, and these are the only degenerate cases which require consideration of V_y^2 and V_z^2.

10. Trilinear forms with $3 \leqq p < q \leqq r$ or $3 < p = q \leqq r$. Although not formulated in the language of trilinear forms the work of Room[17] on the freedom of a projectively generated (determinantal) manifold can be rephrased to give the following theorems:

10.1 THEOREM. *In general, two forms F and G with $3 \leqq p < q \leqq r$ are projectively equivalent if and only if their principal manifolds are projectively equivalent.*

10.2 THEOREM. *In general, two forms F and G with $3 < p = q \leqq r$ are g-equivalent if and only if their principal manifolds are projectively equivalent.*

In a sense 10.1 and 10.2 solve problem I for all cases (p, q, r) not treated in [14] and the earlier sections of this paper. However, the qualifying "in general" leaves much to be desired and requires some clarification.

There are at least two current interpretations of "in general." The modern algebraist means by "general" that the coefficients are independent indeterminants over the base field. The traditional algebraic geometer means by "general" that the coefficients are given elements of the base field but are not in any special relation to each other. In this discussion we shall denote the first interpretation by "generic" and the second by "non-special." The above theorems refer to the second interpretation.

Two completely generic trilinear forms with the same p, q, r are isomorphic but the concept of projective equivalence is meaningless (unless the indeterminants of the two forms are related). On the other hand, projective equivalence always has meaning for non-special forms, but the concept non-special is rather vague and can be made definite for objects of a given category only after the relevant properties of these objects are known.

We can illustrate this by discussing several possible definitions of non-special forms (3, 3, 4). We consider first

10.3 DEFINITION. A form $F(3, 3, 4)$ is said to be *non-special* if

a) V_y^1 is empty.
b) V_y^2 consists of 6 distinct points not on a conic and no 3 on a line.
c) V_y^2 is not projectively equivalent to V_z^2.

With this definition of non-special the following theorem (cf. theorem 6.6) is valid.

[17] [3] Chapter VII, especially p. 124, or [10].

10.4 THEOREM. *In general, $F \sim \bar{F}$ if and only if V_y^2, V_z^2 are projectively equivalent to $\bar{V}_y^2$, $\bar{V}_z^2$, respectively.*

However, 10.4 remains true (cf. 7.2) with a much weaker definition of "in general." We may delete 10.3c and replace 10.3b by

(10.3b′) V_y^2 is zero dimensional of order 6 and W_y is non-nodal.

But this modified definition could hardly have been made without the complete solutions of problems I and II for forms (3, 3, 4) at hand.

The "only if" of 10.4 is valid for any definition of "in general." The "if" follows for any non-degenerate case in which V_y^2 defines a unique W_y. This is the real reason for distinguishing between the non-nodal and nodal cases.

The situation for (3, 3, 4) suggests that the non-special need not be of the strongest possible type and leads to the question: Just how special can a trilinear form (p, q, r) be and still be non-special with respect to theorems 10.1 or 10.2? It is certain that the concept "non-special" must include some generalization of the concepts "non-degenerate" and "nodal," and it is possible that no further requirements need be made.

THEOREMS 10.1 and 10.2 give no clue to the solution of problem II. For forms with $p = q$, Room[18] has a theorem which gives necessary and sufficient conditions that a hypersurface can be a principal manifold, but these conditions being inductive in nature are not completely satisfactory from our point of view. However, in any systematic attempt at solving problem II one would do well to take account of the numerous examples and theorems contained in Room's work.

Appendix. The cases in the table of g-classes are labeled first with a small letter describing V_y^1 and V_z^1 and then a number giving the projective nature of V_z^2. When there are several cases with the same letter and number, subscripts are attached to the letter. The capital letters after the comma describe V_x^1 and are not a part of the label. The key to the letters and numbers used follows:

V_y^1 and V_z^1 are

- *a* both lines
- *b* one a line and one a point
- *c* both points
- *d* one two points and one a single point
- *e* both a single point
- *f* one a single point and one empty
- *g* both empty

V_x^2 is

1. quadric and plane meeting in a conic
2. quadric and plane meeting in two lines
3. cone and plane meeting in a conic
4. cone and plane meeting in two distinct lines
5. cone and plane meeting in a single line

[18] [3] p. 65.

6 three independent planes
7 three coaxial planes
8 two planes one double
9 triple plane
10 whole space
11 ruled cubic with nodal line of first kind[19]
12 ruled cubic with nodal line of second kind[19]

In the description of V_x^1

P single point
P_i i points, $i > 1$
L one line
$\bar{L}_2$ two intersecting lines
L_2 two skew lines
C conic
Q plane
E empty

(In combinations such as LP_i read "line plus i points.")

TABLE OF DEGENERATE g-CLASSES

$$a_1 10,\ \bar{L}_2 \quad \left\| \begin{matrix} y_1 & y_2 & 0 & 0 \\ 0 & 0 & y_3 & 0 \\ 0 & 0 & 0 & y_3 \end{matrix} \right\| \qquad a_2 10,\ L_2 \quad \left\| \begin{matrix} y_1 & y_2 & y_3 & 0 \\ 0 & y_3 & 0 & 0 \\ 0 & 0 & 0 & y_3 \end{matrix} \right\|$$

$$b8,\ Q \quad \left\| \begin{matrix} y_1 & 0 & 0 & y_3 \\ y_2 & 0 & y_3 & 0 \\ 0 & y_3 & 0 & 0 \end{matrix} \right\| \qquad c6,\ LP_2 \quad \left\| \begin{matrix} y_1 & 0 & y_3 & 0 \\ 0 & y_2 & 0 & 0 \\ 0 & 0 & 0 & y_3 \end{matrix} \right\|$$

$$d_1 6,\ LP_2 \quad \left\| \begin{matrix} y_1 & 0 & y_3 & 0 \\ 0 & y_2 & y_3 & 0 \\ 0 & 0 & 0 & y_3 \end{matrix} \right\| \qquad d_2 6,\ LP_1 \quad \left\| \begin{matrix} y_1 & 0 & y_3 & 0 \\ y_3 & y_2 & 0 & 0 \\ 0 & 0 & 0 & y_3 \end{matrix} \right\|$$

$$e9,\ L_2 \quad \left\| \begin{matrix} y_3 & 0 & 0 & 0 \\ y_2 & 0 & 0 & y_3 \\ y_1 & y_2 & y_3 & 0 \end{matrix} \right\| \qquad e_1 8,\ C \quad \left\| \begin{matrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & 0 & y_3 \\ y_1 & 0 & y_2 & 0 \end{matrix} \right\|$$

$$e_2 8,\ L_2 \quad \left\| \begin{matrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & 0 & y_3 \\ y_1 & y_3 & y_2 & 0 \end{matrix} \right\| \qquad e_3 8,\ C \quad \left\| \begin{matrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & y_3 & 0 \\ y_1 & 0 & y_2 & y_3 \end{matrix} \right\|$$

$$e_4 8,\ LP \quad \left\| \begin{matrix} y_3 & 0 & 0 & 0 \\ y_2 & 0 & y_3 & 0 \\ 0 & y_1 & 0 & y_2 \end{matrix} \right\| \qquad e_5 8,\ L \quad \left\| \begin{matrix} y_3 & 0 & 0 & 0 \\ y_2 & y_3 & 0 & 0 \\ 0 & y_1 & y_2 & y_3 \end{matrix} \right\|$$

$$e_6 8,\ L \quad \left\| \begin{matrix} y_3 & 0 & 0 & 0 \\ y_2 & 0 & y_3 & 0 \\ 0 & y_1 & y_2 & y_3 \end{matrix} \right\| \qquad e_7 7,\ P_2 \quad \left\| \begin{matrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & y_3 & 0 \\ y_1 & y_1 & 0 & y_2 \end{matrix} \right\|$$

[19] See [1], p. 61.

$$e_27,\ P \quad \begin{Vmatrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & y_3 & 0 \\ y_1 & y_1 & y_2 & y_3 \end{Vmatrix}$$

$$e_26,\ P_4 \quad \begin{Vmatrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & 0 & y_3 \\ 0 & y_3 & y_1 & y_2 \end{Vmatrix}$$

$$e_46,\ LP \quad \begin{Vmatrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & 0 & y_3 \\ y_2 & 0 & y_1 & 0 \end{Vmatrix}$$

$$e_66,\ P_2 \quad \begin{Vmatrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & 0 & y_3 \\ y_2 & y_3 & y_1 & 0 \end{Vmatrix}$$

$$e2,\ L_2 \quad \begin{Vmatrix} y_1 & 0 & 0 & 0 \\ 0 & y_2 & y_3 & 0 \\ y_3 & 0 & 0 & y_2 \end{Vmatrix}$$

$$f_15,\ P_2 \quad \begin{Vmatrix} y_1 & 0 & 0 & y_2 \\ y_2 & 0 & y_3 & 0 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_35,\ P \quad \begin{Vmatrix} y_1 & y_2 & 0 & y_3 \\ y_2 & 0 & y_3 & 0 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_24,\ P_3 \quad \begin{Vmatrix} y_1 & 0 & 0 & y_2 + y_3 \\ y_2 & 0 & y_3 & 0 \\ y_3 & y_2 & 0 & 0 \end{Vmatrix}$$

$$f_13,\ P_2 \quad \begin{Vmatrix} y_1 & 0 & 0 & y_2 \\ 0 & y_2 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_33,\ P_4 \quad \begin{Vmatrix} y_1 & y_2 + y_3 & 0 & 0 \\ 0 & y_2 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_22,\ P_3 \quad \begin{Vmatrix} y_1 & y_2 & y_3 & 0 \\ y_2 & 0 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_42,\ L \quad \begin{Vmatrix} y_1 & 0 & y_3 & 0 \\ y_2 & 0 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_62,\ LP \quad \begin{Vmatrix} y_1 & 0 & y_2 & 0 \\ y_2 & 0 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$e_16,\ P_3 \quad \begin{Vmatrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & 0 & y_3 \\ 0 & 0 & y_1 & y_2 \end{Vmatrix}$$

$$e_36,\ P_2 \quad \begin{Vmatrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & y_3 & 0 \\ y_2 & 0 & y_1 & y_3 \end{Vmatrix}$$

$$e_56,\ P_3 \quad \begin{Vmatrix} y_3 & 0 & 0 & 0 \\ 0 & y_2 & 0 & y_3 \\ y_2 & 0 & y_1 & y_3 \end{Vmatrix}$$

$$e3,\ CP \quad \begin{Vmatrix} y_1 & 0 & 0 & 0 \\ 0 & y_2 & y_3 & 0 \\ 0 & 0 & y_2 & y_3 \end{Vmatrix}$$

$$e1,\ C \quad \begin{Vmatrix} y_1 & 0 & 0 & 0 \\ 0 & y_2 & y_3 & 0 \\ y_2 & 0 & y_2 & y_3 \end{Vmatrix}$$

$$f_25,\ L \quad \begin{Vmatrix} y_1 & 0 & 0 & y_3 \\ y_2 & 0 & y_3 & 0 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_14,\ P_2 \quad \begin{Vmatrix} y_1 & y_3 & 0 & y_2 \\ y_2 & 0 & y_3 & 0 \\ y_3 & y_2 & 0 & 0 \end{Vmatrix}$$

$$f_34,\ LP \quad \begin{Vmatrix} y_1 & 0 & 0 & y_2 \\ y_2 & 0 & y_3 & 0 \\ y_3 & y_2 & 0 & 0 \end{Vmatrix}$$

$$f_23,\ P_3 \quad \begin{Vmatrix} y_1 & y_2 & 0 & 0 \\ 0 & y_2 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_12,\ P_2 \quad \begin{Vmatrix} y_1 & y_2 & 0 & 0 \\ y_2 & 0 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_32,\ L \quad \begin{Vmatrix} y_1 & 0 & 0 & y_2 \\ y_2 & 0 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_52,\ P \quad \begin{Vmatrix} y_1 & 0 & y_3 & y_2 \\ y_2 & 0 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_72,\ P_2 \quad \begin{Vmatrix} y_1 & 0 & y_2 & y_2 \\ y_2 & 0 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{Vmatrix}$$

$$f_1 1,\ P \quad \left\| \begin{matrix} y_1 & 0 & 0 & y_2 \\ y_2 & y_2 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{matrix} \right\| \qquad f_2 1,\ P_2 \quad \left\| \begin{matrix} y_1 & y_2 & 0 & 0 \\ y_2 & y_2 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{matrix} \right\|$$

$$f_3 1,\ P_2 \quad \left\| \begin{matrix} y_1 & y_2 & y_2 & 0 \\ y_2 & y_2 & 0 & y_3 \\ 0 & y_3 & y_2 & 0 \end{matrix} \right\| \qquad g12,\ L \quad \left\| \begin{matrix} y_1 & y_2 & y_3 & 0 \\ 0 & 0 & y_2 & y_1 \\ 0 & 0 & y_1 & y_3 \end{matrix} \right\|$$

$$g11,\ L \quad \left\| \begin{matrix} y_1 & y_2 & 0 & 0 \\ 0 & 0 & y_2 & y_3 \\ 0 & 0 & y_3 & y_1 \end{matrix} \right\| \qquad g8,\ P \quad \left\| \begin{matrix} 0 & y_3 & -y_2 & y_1 \\ -y_3 & 0 & y_1 & 0 \\ y_2 & -y_1 & 0 & 0 \end{matrix} \right\|$$

$$g6,\ P \quad \left\| \begin{matrix} 0 & y_3 & -y_2 & 0 \\ -y_3 & 0 & y_1 & y_1 \\ y_2 & -y_1 & 0 & 0 \end{matrix} \right\| \qquad g3,\ P_2 \quad \left\| \begin{matrix} y_2 & y_3 & 0 & 0 \\ 0 & 0 & y_2 & y_1 \\ 0 & y_1 & y_3 & 0 \end{matrix} \right\|$$

$$g_1 2,\ P \quad \left\| \begin{matrix} y_2 & y_3 & 0 & y_1 \\ 0 & ay_1 & y_2 & 0 \\ y_1 & 0 & y_3 & 0 \end{matrix} \right\| \qquad g_2 2,\ P_2 \quad \left\| \begin{matrix} y_2 & y_3 & y_1 & 0 \\ 0 & 0 & y_2 & y_1 \\ y_1 & 0 & y_3 & 0 \end{matrix} \right\|$$

$$g_3 2,\ P_2 \quad \left\| \begin{matrix} y_2 & y_3 & 0 & 0 \\ 0 & 0 & y_2 & y_1 \\ y_1 & 0 & y_3 + y_1 & 0 \end{matrix} \right\| \qquad g_1 1,\ E \quad \left\| \begin{matrix} 0 & y_3 & -y_2 & y_1 \\ -y_3 & 0 & y_1 & 0 \\ y_2 & -y_1 & 0 & -y_2 \end{matrix} \right\|$$

$$g_2 1,\ P_2. \quad \left\| \begin{matrix} y_2 & y_3 & y_1 & 0 \\ 0 & 0 & y_2 & y_1 \\ 0 & y_1 & y_3 & 0 \end{matrix} \right\|$$

UNIVERSITY OF MICHIGAN.

BIBLIOGRAPHY

(1) A. B. Bassett, *A Treatise on the Geometry of Surfaces*, Cambridge, 1910.

(2) Julian L. Coolidge, *Algebraic Plane Curves*, Oxford, 1931.

(3) T. G. Room, *The Geometry of Determinantal Loci*, Cambridge, 1938.

(4) Severi-Löffler, *Vorlesungen über Algebraische Geometrie*, Teubner, 1921.

(5) B. L. Van der Waerden, *Moderne Algebra*, Springer, 1931.

(6) Josephine Chanler, *The Invariant Theory of the Ternary Trilinear Form*, Duke Mathematical Journal, vol. 5 (1939), pp. 552–566.

(7) A. B. Coble, *Point Sets and Allied Cremona Groups, Part I*, Transactions of the American Mathematical Society, vol. 16 (1915), pp. 155–198.

(8) L. E. Dickson, *Determination of all General Homogeneous Polynomials Expressible as Determinants with Linear Elements*, Transactions of the American Mathematical Society, vol. 22 (1921), pp. 167–179.

(9) L. E. Dickson, *Singular Case of Pairs of Bilinear, Quadratic, or Hermitian Forms*, Transactions of the American Mathematical Society, vol. 29 (1927), pp. 239–253.

(10) T. G. Room, *Freedoms of Determinantal Manifolds*, Proceedings of the London Mathematical Society, (2) vol. 36 (1934), pp. 1–28.

(11) Virgil Snyder, *Room on Determinantal Loci*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 499–501.

(12) R. M. Thrall, *Apolarity of Trilinear Forms and Pencils of Bilinear Forms*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 678–683.

(13) R. M. Thrall, *Metabelian Groups and Trilinear Forms*, American Journal of Mathematics, vol. 60 (1937), pp. 383–415.

(14) R. M. Thrall and Josephine Chanler, *Ternary Trilinear Forms in the Field of Complex Numbers*, Duke Mathematical Journal, vol. 4 (1938), pp. 678–690.

Annals of Mathematics
Vol. 42, No. 2, April, 1941

HOMOMORPHISM OF GROUPS

By J. H. M. Wedderburn

(Received October 18, 1940)

As usually given homomorphism between groups is a many-one relation but it will be shown below that the treatment of the many-many relation is just as simple and in some respects clearer because of the symmetry.

Let G and H be two groups each divided into mutually exclusive sets

$$G = G_1 + G_2 + \cdots = \sum G_p, \qquad H = H_1 + H_2 + \cdots = \sum H_p$$
$$G_p \frown G_q = 0, \qquad H_p \frown H_q = 0 \qquad (p \neq q)$$

where the notation is not to be taken to mean that the sets are denumerable; the elements of a set G_p will be denoted by $g_p, g'_p, \cdots$. Then G is said to be homomorphic to H, $G \sim H$, if

$$g_p g_q = g_r \supset H_p H_q \leqq H_r. \tag{1}$$

If $g'_p g'_q = g_s$, then $H_p H_q \leqq H_s$; but $H_r \frown H_s = 0$ if $r \neq s$; hence $r = s$, that is, (1) implies

$$G_p G_q \leqq G_r. \tag{2}$$

Again, if $h_a h_b = h_c$, let $g_a g_b = g_d$; then $H_a H_b \leqq H_d$. But $H_a H_b \geqq h_c < H_c$; therefore $H_c \frown H_d \neq 0$ so that $d = c$ and, since $g_d < G_d$ no matter what elements g_a and g_b are in G_a and G_b, we have

$$h_a h_b = h_c \supset G_a G_b \leqq G_c$$

that is, the relation of homomorphism is reflexive.

Suppose now that the identity g_1 is in G_1; then

$$g_1 g_p = g_p = g_p g_1 \supset G_1 G_p = G_p = G_p G_1$$

for all p, and therefore

$$H_1 H_p \leqq H_p, \qquad H_p H_1 \leqq H_p. \tag{3}$$

If the identity of H lies in H_a, it follows similarly that

$$H_a H_p = H_p = H_p H_a$$

and in particular

$$H_a H_1 = H_1 = H_1 H_a.$$

But from (3) $H_a H_1 \leqq H_a$ and hence $a = 1$, so that the identity of H lies in H_1.

Since G is a group, any g'_1 has an inverse, say g_p, such that

$$g'_1 g_p = g_1 = g_p g'_1.$$

Hence $G_1G_p \leqq G_1$: but $G_1G_p = G_p$ and hence $G_p = G_1$, that is, G_1 is a group. Since the relation is reflexive, H_1 is also a group. Let g_p be any element of G_p $(p \neq 1)$ and let $g_p^{-1} = g_q < G_q$; then $G_pG_q \frown G_1 \geqq g_1$ and therefore from (2) $G_pG_q \leqq G_1$. But $G_pg_q \leqq G_pG_q$ so that $G_pg_q \leqq G_1$. But $G_pg_q \leqq G_pG_q$ so that $G_pg_q \leqq G_1$; hence

$$G_p = G_pg_qg_p \leqq G_1g_p \leqq G_p$$

so that

$$G_p = G_1g_p \tag{4}$$

for every element g_p in G_p. Similarly $H_p = H_1h_p$ and $G_p = g_pG_1$, $H_p = h_pH_1$, so that G_1 and H_1 are invariant in G and H. The final result can now be stated.

THEOREM. *If $G \sim H$, then also $H \sim G$. If G_1 contains the identity of G, then $H_1 \sim G_1$ contains the identity of H. Further G_1 and H_1 are invariant subgroups of G and H, respectively, and $G/G_1 \simeq H/H_1$.*

Let G_1 be a subgroup of G minimal with respect to the property that there is a homomorphism with H given by $G/G_1 \simeq H/H_1$; and let G_2 be a second such subgroup. If we set

$$G = \sum G_1g_i = \sum G_2g_i', \qquad H = \sum H_1h_i = \sum H_2h_i'$$

then $g_i \sim h_i$ in the first homomorphism, and $g_i' \sim h_i'$ in the second. Let

$$B = G_1 \frown G_2, \qquad G_1 = \sum B\gamma_i, \qquad G_2 = \sum B\gamma_i'$$

$$C = H_1 \frown H_2, \qquad H_1 = \sum C\eta_i, \qquad H_2 = \sum C\eta_i' ;$$

then $G_1G_2 = \sum B\gamma_i\gamma_j'$, and

$$G = \sum G_1G_2\gamma_k'' = \sum B\gamma_i\gamma_j'\gamma_k''$$

$$H = \sum H_1H_2\eta_k'' = \sum C\eta_i\eta_j'\eta_k''$$

which gives a homomorphism between G and H by means of B and C. But B is a subgroup of G_1 which is minimal and hence $B = G_1$ and so $G_2 = G_1$. The minimal subgroup G_1 is therefore unique.

PRINCETON, NEW JERSEY.

Annals of Mathematics
Vol. 42, No. 2, April, 1941

AN ANALOGUE TO MINKOWSKI'S GEOMETRY OF NUMBERS IN A FIELD OF SERIES

By Kurt Mahler

(Received November 16, 1939)

Minkowski, in his "Geometrie der Zahlen" (Leipzig 1910), studied properties of a convex body in a space R_n of n dimensions with respect to the set of all lattice points. Let $F(X) = F(x_1, \cdots, x_n)$ be a distance function, i.e. a function satisfying the conditions

$$F(0) = 0, F(X) > 0 \text{ if } X \neq 0;$$
$$F(tX) = |t| F(X) \text{ for all real } t;$$
$$F(X - Y) \leqq F(X) + F(Y).$$

The inequality $F(X) \leqq 1$ defines a convex body in R_n which has its centre at the origin $X = 0$. Suppose that this body has the volume V. The well known result of Minkowski asserts that *if $V \geqq 2^n$, then the body contains at least one (and so at least two) lattice points different from* 0. This theorem is contained in the following deeper result of Minkowski (G.d.Z. §§50–53): "*There are n independent lattice points $X^{(1)}, X^{(2)}, \cdots, X^{(n)}$ in R_n with the following properties*: (1) $F(X^{(1)}) = \sigma^{(1)}$ *is the minimum of $F(X)$ in all lattice points $X \neq 0$, and for $k \geqq 2$, $F(X^{(k)}) = \sigma^{(k)}$ is the minimum of $F(X)$ in all lattice points X which are independent of $X^{(1)}, \cdots, X^{(k-1)}$.* (2) *The determinant D of the points $X^{(1)}, \cdots, X^{(n)}$ satisfies the inequalities*

$$1 \leqq |D| \leqq n!.$$

(3) *The numbers $\sigma^{(k)}$ depend only on $F(X)$ and not on the special choice of the lattice points $X^{(k)}$, and they satisfy the inequalities*

$$0 < \sigma^{(1)} \leqq \sigma^{(2)} \leqq \cdots \leqq \sigma^{(n)}, \qquad \frac{2^n}{n!} \leqq V\sigma^{(1)}\sigma^{(2)} \cdots \sigma^{(n)} \leqq 2^n."$$

(A new simple proof for the last part of this theorem was given by *H. Davenport*, Quart. Journ. Math. (Oxford Ser.), Vol. 10 (1939), 119–121).

From Minkowski's theorem, properties of general classes of convex bodies can be obtained. For instance, there is a *convex body $G(Y) \leqq 1$ polar to $F(X) \leqq 1$*, and to this body correspond by the theorem n minima $\tau^{(1)}, \tau^{(2)}, \cdots, \tau^{(n)}$. I have proved (Časopis 68 (1939), 93–102), that *these minima are related to the σ's by the inequalities*

$$1 \leqq \sigma^{(h)}\tau^{(n-h+1)} \leqq (n!)^2 \qquad (h = 1, 2, \cdots, n).$$

From this result, applications to inhomogeneous Diophantine inequalities can be made, and in particular, generalizations of *Kronecker's* theorem can be obtained.

The present paper does *not* deal with ordinary convex bodies in a real space. The n-dimensional space P_n with which we shall be concerned has its coordinates in a field $\mathfrak{K}$ with a non-*Archimedean* valuation $|x|$; a distance function is any function satisfying

$$F(0) = 0, F(X) > 0 \text{ if } X \neq 0,$$
$$F(tX) = |t| F(X) \text{ for all } t \text{ in } \mathfrak{K},$$
$$F(X - Y) \leqq \max (F(X), F(Y)).$$

The inequality $F(X) \leqq \tau$ then defines the convex body $C(\tau)$, if $\tau > 0$. We show that every convex body is bounded, and that it has properties similar to a parallelepiped in real space.

In particular, let $\mathfrak{K}$ be the field of all Laurent series

$$x = \alpha_f z^f + \alpha_{f-1} z^{f-1} + \alpha_{f-2} z^{f-2} + \cdots$$

with coefficients in an arbitrary field $\mathfrak{k}$; the valuation $|x|$ is defined as $|0| = 0$, and $|x| = e^f$ if $\alpha_f \neq 0$. Further let Λ_n be the modul of all points in P_n, the coordinates of which are polynomials in z with coefficients in $\mathfrak{k}$; these points we call *lattice points*. We consider only distance functions $F(X)$ which for all $X \neq 0$ in P_n are always as integral power of e. We shall define a certain positive constant V as the volume of $C(1)$; this constant is invariant under all linear transformations of P_n with determinant 1, and the volume of $C(1)$ and that of its polar reciprocal body $C'(1)$ have the product 1. In analogy to Minkowski's theorem, the following theorem holds: "*There are n independent lattice points $X^{(1)}, \cdots, X^{(n)}$ in P_n with the following properties*: 1) *$F(X^{(1)})$ is the minimum of $F(X)$ in all lattice points $X \neq 0$, and for $k \geqq 2$, $F(X^{(k)})$ is the minimum of $F(X)$ in all lattice points X which are independent of $X^{(1)}, \cdots, X^{(k-1)}$.* 2) *The determinant of the points $X^{(1)}, \cdots, X^{(n)}$ is* 1. 3) *The numbers $F(X^{(k)}) = \sigma^{(k)}$, which depend only on $F(X)$ and not on the special choice of the lattice points X, satisfy the formulae*

$$0 < \sigma^{(1)} \leqq \sigma^{(2)} \leqq \cdots \leqq \sigma^{(n)}, \qquad \sigma^{(1)} \sigma^{(2)} \cdots \sigma^{(n)} = \frac{1}{V}."$$

Further, we have similar minima $\tau^{(1)}, \cdots, \tau^{(n)}$ for the distance function $G(Y)$ which defines the polar body $C'(1)$; these are related with the σ's by the equations

$$\sigma^{(h)} \tau^{(n-h+1)} = 1 \qquad (h = 1, 2, \cdots, n).$$

These two results can be used to study special Diophantine problems in P_n; a few of them are considered as examples. All the proofs in this paper are based on the methods of Minkowski, and in one final paragraph I make use of ideas of C. L. Siegel.

I. Convex Domains in Non-Archimedean Spaces

1. Notation. In this chapter, we denote by

$\mathfrak{R}$ *an arbitrary field,*

$|x|$ *a non-Archimedean valuation of the elements x of $\mathfrak{R}$,*[1]

$\mathfrak{K}$ *the perfect extension of $\mathfrak{R}$ with respect to this valuation,*

P_n *the n-dimensional space of all points or vectors*

$$X = (x_1, \cdots, x_n),$$

where the coordinates $x_1, \cdots, x_n$ lie in $\mathfrak{K}$,

$|X|$ *the length of the vector X, viz.*

$$|X| = \max(|x_1|, \cdots, |x_n|).$$

We apply the usual notation for vectors in P_n; thus if

$$X = (x_1, \cdots, x_n) \quad \text{and} \quad Y = (y_1, \cdots, y_n),$$

and a belongs to $\mathfrak{K}$, then we write

$$X \mp Y = (x_1 \mp y_1, \cdots, x_n \mp y_n),$$
$$aX = (ax_1, \cdots, ax_n),$$
$$XY = \sum_{h=1}^{n} x_h y_h.$$

For instance, the length $|X|$ of X has the properties:

(1) $|X| \geqq 0$, with equality if and only if $X = (0, \cdots, 0) = 0$;

(2) $|aX| = |a||X|$, if a is any element of $\mathfrak{K}$;

(3) $|X \mp Y| \leqq \max(|X|, |Y|)$;

(4) $|XY| \leqq |X||Y|$.

If $\mathfrak{D}$ is any sub-ring of $\mathfrak{K}$, and $X^{(1)}, \cdots, X^{(r)}$ are vectors in P_n, then these are called $\mathfrak{D}$-dependent, or $\mathfrak{D}$-independent, according as there exist, or do not exist elements $a_1, \cdots, a_r$ of $\mathfrak{D}$ not all zero, such that

$$a_1X^{(1)} + \cdots + a_rX^{(r)} = 0.$$

A set of vectors of P_n is called a $\mathfrak{D}$-modul, if with X and Y it also contains $aX + bY$, where a and b are arbitrary elements of $\mathfrak{D}$; the modul has the dimen-

[1] This means that the function $|x|$ satisfies the conditions:

$$|0| = 0, \text{ but } |x| > 0 \text{ for } x \neq 0,$$
$$|xy| = |x||y|,$$
$$|x \mp y| \leqq \max(|x|, |y|).$$

sion m, if there are m, but not $m + 1$, $\mathfrak{D}$-independent elements in it. The dimension of a $\mathfrak{K}$-modul is at most n, while that of any other class of moduli need not be finite.

2. The distance function $F(X)$. A function $F(X)$ of the variable point X in P_n is called a general distance function, if it has the properties:

(A): $$F(X) \geqq 0;$$

(B): $$F(aX) = |a| F(X) \text{ for all } a \text{ in } \mathfrak{K}, \text{ hence } F(0) = 0;$$

(C): $$F(X \mp Y) \leqq \max (F(X), F(Y));$$

it is called a special distance function or simply a distance function, if instead of (A) it satisfies the stronger condition

(A′): $$F(X) > 0 \text{ for } X \neq 0.$$

If τ is a positive number, then the set $C(\tau)$ of all points X with

$$F(X) \leqq \tau$$

is called a convex set;[2] if $F(X)$ is a special distance function, then it is called a convex body. It is clear from the definition of $F(X)$ that a convex set $C(\tau)$ contains the origin 0, and that with X and Y also $aX + bY$ belong to it, if a and b are elements of $\mathfrak{K}$ such that $|a| \leqq 1$, $|b| \leqq 1$. Further, if

$$E^{(1)} = (1, 0, \cdots, 0), E^{(2)} = (0, 1, \cdots, 0), \cdots, E^{(n)} = (0, 0, \cdots, 1)$$

are the n unit vectors of the coordinate system, then

$$X = x_1 E^{(1)} + \cdots + x_n E^{(n)}, \quad \text{i.e.} \quad F(X) \leqq \max_{h=1,2,\cdots,n} (|x_h| F(E^{(h)})),$$

and therefore

(5) $$F(X) \leqq \Gamma |X|,$$

where Γ is the positive constant

$$\Gamma = \max_{h=1,2,\cdots,n} (F(E^{(h)})).$$

$C(\tau)$ contains therefore all points of the cube

$$|X| \leqq \frac{\tau}{\Gamma}.$$

We prove now that *for special distance functions there is a* second *positive constant* γ, *such that for all points in* P_n

(6) $$F(X) \geqq \gamma |X|.$$

[2] We consider only convex sets and bodies as defined; they are obviously symmetrical with respect to the origin.

PROOF: We assume that (6) is not true and show that this leads to a contradiction.

By hypothesis, there is an infinite sequence S of points

$$X^{(h)} = (x_1^{(h)}, \cdots, x_n^{(h)}) \neq 0 \qquad (h = 1, 2, 3, \cdots),$$

such that

$$\lim_{h\to\infty} \frac{F(X^{(h)})}{|X^{(h)}|} = 0.$$

Since

$$\frac{F(aX)}{|aX|} = \frac{F(X)}{|X|}$$

for all $a \neq 0$ in $\mathfrak{K}$, we may assume that for the elements of S

$$\lim_{h\to\infty} F(X^{(h)}) = 0, \qquad |X^{(h)}| = 1,$$

so that in particular the n real sequences

$$|x_k^{(1)}|, |x_k^{(2)}|, |x_k^{(2)}|, \cdots \qquad (k = 1, 2, \cdots, n)$$

are bounded.

Hence we can replace S by an infinite sub-sequence which we again call S: $X^{(1)}, X^{(2)}, X^{(3)}, \cdots$, such that the n real limits

$$a_k = \lim_{h\to\infty} |x_k^{(h)}| \qquad (k = 1, 2, \cdots, n) \tag{7}$$

exist and satisfy the equation

$$\max_{k=1,2,\cdots,n} a_k = 1.$$

We call S a sequence of rank m, if exactly m of the limits $a_1, a_2, \cdots, a_n$ do not vanish; without loss of generality, these are the m first limits $a_1, a_2, \cdots, a_m$. Obviously $1 \leqq m \leqq n$.

If the rank $m = 1$, then for large h

$$|x_1^{(h)}| = 1, \quad \text{and} \quad \frac{X^{(h)}}{x_1^{(h)}} = \left(1, \frac{x_2^{(h)}}{x_1^{(h)}}, \cdots, \frac{x_n^{(h)}}{x_1^{(h)}}\right), = E^{(1)} + X^{*(h)}$$

say, where

$$\lim_{h\to\infty} |X^{*(h)}| = 0.$$

Hence by (5)

$$F(E^{(1)}) = F\left(\frac{X^{(h)}}{x_1^{(h)}} - X^{*(h)}\right) \leqq \max\left(\frac{F(X^{(h)})}{|x_1^{(h)}|}, F(X^{*(h)})\right)$$

$$\leqq \max\,(F(X^{(h)}), \Gamma\,|X^{*(h)}|),$$

and therefore for $h \to \infty$

$$0 \leqq F(E^{(1)}) \leqq 0, \text{ i.e. } F(E^{(1)}) = 0,$$

which is not true.

Hence the rank $m \geqq 2$. Put

$$X^{(g,h)} = \frac{X^{(g)}}{x_m^{(g)}} - \frac{X^{(h)}}{x_m^{(h)}} = (x_1^{(g,h)}, \cdots, x_m^{(g,h)}).$$

Then from (7) for large g, h

$$F(X^{(g,h)} \leqq \max\left(\frac{F(X^{(g)})}{|x_m^{(g)}|}, \frac{F(X^{(h)})}{|x_m^{(h)}|}\right) \leqq \frac{2}{a_m} \max(F(X^{(g)}), F(X^{(h)})),$$

and therefore

$$\lim_{\substack{g\to\infty \\ h\to\infty}} F(X^{(g,h)}) = 0.$$

Two cases are now possible:

a: The limit

$$\lim_{\substack{g\to\infty \\ h\to\infty}} |X^{(g,h)}| = \lim_{\substack{g\to\infty \\ h\to\infty}} \max(|x_1^{(g,h)}|, \cdots, |x_n^{(g,h)}|)$$

exists and is zero. Hence the n limits in $\mathfrak{K}$

(8) $$x_k^* = \lim_{h\to\infty} \frac{x_k^{(h)}}{x_m^{(h)}} \qquad (k = 1, 2, \cdots, n)$$

all exist, and in particular

$$x_m^* = \lim_{h\to\infty} 1 = 1,$$

so that

$$X^* = (x_1^*, \cdots, x_n^*) \neq 0.$$

By the continuity of $F(X)$,[3]

$$F(X^*) = \lim_{h\to\infty} F\left(\frac{X^{(h)}}{x_m^{(h)}}\right) = \frac{1}{a_m} \lim_{h\to\infty} F(X^{(h)}) = 0,$$

which is not true.

b: The limit

$$\lim_{\substack{g\to\infty \\ h\to\infty}} |X^{(g,h)}|$$

---

[3] If $\epsilon > 0$ is given, then there is a $\delta > 0$, such that $|F(X) - F(Y)| < \epsilon$ for $|X - Y| < \delta$, as follows easily from the properties (B), (C), and (5).

either does not exist, or exists and is different from zero. That implies that at least one of the limits (8) does not exist. Now obviously

$$\lim_{\substack{g\to\infty \\ h\to\infty}} | x_k^{(g,h)} | = 0 \qquad (k = m, m+1, \cdots, n),$$

since for large g, h

$$x_m^{(g,h)} = 0; \qquad | x_k^{(g,h)} | = \left| \frac{x_k^{(g)}}{x_m^{(g)}} - \frac{x_k^{(h)}}{x_m^{(h)}} \right| \leqq \frac{2}{a_m} \max (| x_k^{(g)} |, | x_k^{(h)} |)$$

$$(k = m+1, \cdots, n).$$

Hence the index μ of this non-existing limit (8) is $\leqq m - 1$. For this index,

$$\lim_{\substack{g\to\infty \\ h\to\infty}} x_\mu^{(g,h)}$$

either does not exist or exists and is different from zero. Hence there is an infinite one-dimensional sub-sequence

$$X^{(g_i,h_i)} \qquad (i = 1, 2, 3, \cdots) \tag{9}$$

of the double sequence $X^{(g,h)}$, such that for all i

$$| x_\mu^{(g_i,h_i)} | \geqq c,$$

where c is a positive constant. Further obviously

$$\lim_{i\to\infty} F(X^{(g_i,h_i)}) = 0,$$

$$\lim_{i\to\infty} | x_k^{(g_i,h_i)} | = 0 \qquad (k = m, m+1, \cdots, n),$$

and all $m - 1$ first coordinates

$$x_k^{(g_i,h_i)} \qquad (k = 1, 2, \cdots, m-1)$$

are bounded for $i \to \infty$.

Let ξ_i, for every i, be the coordinate

$$x_k^{(g_i,h_i)} \qquad (k = 1, 2, \cdots, m-1)$$

of maximum value $| x_k^{(g_i,h_i)} |$; hence

$$| \xi_i | \geqq c, \qquad \text{since} \qquad | \xi_i | \geqq | x_\mu^{(g_i,h_i)} |.$$

Then there is an infinite subsequence

$$X^{(g_{i_j},h_{i_j})} \qquad (j = 1, 2, 3, \cdots)$$

of the sequence (9), such that, if

$$X'^{(j)} = \frac{X^{(g_{i_j},h_{i_j})}}{\xi_{i_j}} = (x_1'^{(j)}, \cdots, x_n'^{(j)}) \qquad (j = 1, 2, 3, \cdots),$$

then all n limits

$$\lim_{j\to\infty} |x_k'^{(j)}| = a_k' \qquad (k = 1, 2, \cdots, n)$$

exist and satisfy the equations

$$\max(a_1', \cdots, a_n') = 1, \qquad a_m = a_{m+1} = \cdots = a_n = 0,$$

and

$$0 \leqq \lim_{j\to\infty} F(X'^{(j)}) \leqq \frac{1}{C} \lim_{j\to\infty} F(X^{(g_i, h_i)}) = 0, \quad \text{i.e.} \quad \lim_{j\to\infty} F(X'^{(j)}) = 0,$$

Therefore the new sequence S'

$$X'^{(1)}, X'^{(2)}, X'^{(3)}, \cdots$$

has the same properties as S, but is of lower rank. Hence by induction with respect to the rank, a contradiction follows also in this case.—

By the inequality (6), all points of the convex body $C(\tau)$ lie in the finite cube

$$|X| \leqq \frac{\tau}{\gamma};$$

a convex body is therefore bounded. Conversely, if a convex set is bounded, then it is a convex body. For if its distance function $F(X)$ is not special, then there is at least one point $X^{(0)} \neq 0$, such that $F(X^{(0)}) = 0$; hence all points of the straight line passing through $X^{(0)}$ and the origin 0 belong to the set.

3. The character of a convex body. Let $C(\tau)$ be a convex body, $F(X)$ its distance function. If $X' \neq 0$ is an arbitrary vector, then the point $X = aX'$, where a is an element of $\mathfrak{K}$, lies in $C(\tau)$ provided that $|a|$ is either sufficiently small and positive, or 0. Hence for every index $h = 1, 2, \cdots, n$, the set S_h of all points

$$X = (x_1, \cdots, x_n) \quad \text{with} \quad x_1 = \cdots = x_{h-1} = 0, \quad x_h \neq 0$$

of $C(\tau)$ is not empty and contains an infinity of elements. By (6),

$$|x_h| \leqq \frac{\tau}{\gamma}$$

for the points of S_h. Therefore $|x_h|$ has a positive upper bound ξ_h in this set, and to every $\epsilon > 0$ there is a point

$$X_\epsilon^{(h)} = (x_{1\epsilon}^{(h)}, \cdots, x_{n\epsilon}^{(h)}),$$

for which

$$F(X_\epsilon^{(h)}) \leqq \tau, \qquad x_{1\epsilon}^{(h)} = \cdots = x_{h-1\,\epsilon}^{(h)} = 0, \qquad \frac{\xi_h}{1+\epsilon} < |x_{h\epsilon}^{(h)}| \leqq \xi_h,$$

whereas there is no point X for which

$$F(X) \leqq \tau, \qquad x_1 = \cdots = x_{h-1} = 0, \qquad |x_h| > \xi_h .$$

The system of the n points

$$X_\epsilon^{(1)}, X_\epsilon^{(2)}, \cdots, X_\epsilon^{(n)}$$

corresponding to ϵ is obviously $\mathfrak{K}$-independent, and any point X of P_n can be written as

$$X = u_{1\epsilon}X_\epsilon^{(1)} + \cdots + u_{n\epsilon}X_\epsilon^{(n)},$$

where the u's belong to $\mathfrak{K}$ and are given explicitly by

$$u_{h\epsilon} = \sum_{k=1}^{n} \alpha_{hk\epsilon} x_k \qquad (h = 1, 2, \cdots, n)$$

with a matrix

$$(\alpha_{hk\epsilon})_{h,k=1,2,\cdots,n}$$

of non-vanishing determinant and elements depending on ϵ, but not on X.

We distinguish now whether the valuation $|x|$ of $\mathfrak{K}$, is *discrete* or not.

If $|x|$ is discrete, then there is a constant $b > 1$, such that for all $x \neq 0$ in $\mathfrak{K}$[4]

$$|x| = b^g$$

---

[4] If $|x|$ is discreet, then $F(X)$ has a similar property: *The set s of its values for X in P_n has no point of accumulation except* 0. This is clear for $n = 1$, for then all vectors are multipla of the unit vector (1). Suppose that the statement has already been proved for all spaces of $n - 1$ dimensions, but that it is not true in P_n. There is therefore an infinite sequence Σ of points

$$X^{(k)} = (x_1^{(k)}, \cdots, x_n^{(k)}) \qquad (k = 1, 2, 3, \cdots)$$

in P_n, such that all numbers

$$F(X^{(1)}), \qquad F(X^{(2)}), \qquad F(X^{(3)}), \cdots$$

are different from each other, and that the limit

$$\lim_{k\to\infty} F(X^{(k)}) = \lambda$$

exists and is positive. Write

$$X^{(k)} = x_1^{(k)} E^{(1)} + X^{(k)*} \qquad (k = 1, 2, 3, \cdots)$$

where

$$X^{(k)*} = (0, x_2^{(k)}, \cdots, x_n^{(k)}) \qquad (k = 1, 2, 3, \cdots)$$

lies in the $(n - 1)$-dimensional subspace P_{n-1}: $x_1 = 0$, of P_n. By (6), $|x_1^{(k)}|$ is bounded in Σ; hence we may assume that

$$\lim_{k\to\infty} |x_1^{(k)}| = \mu,$$

with a rational integer g depending on x. In this case the set of values $| x_{h\epsilon}^{(h)} |$ satisfies the equations

$$| x_{h\epsilon}^{(h)} | = \xi_h \qquad (h = 1, 2, \cdots, n)$$

for all sufficiently small ϵ. We assume that ϵ is sufficiently small and omit the index ϵ. Put

$$\Phi_\tau(X) = \tau \max (|u_1|, \cdots, |u_n|) = \tau \max_{h=1,2,\cdots,n} \left(\left| \sum_{k=1}^{n} \alpha_{hk} x_k \right| \right).$$

Then obviously

$$F(X) \leqq \tau, \qquad \text{if} \qquad \Phi_\tau(X) \leqq \tau.$$

Conversely let X be any point in $C(\tau)$. Then

$$| x_1 | \leqq \xi_1$$

and therefore

$$|u_1| = \frac{|x_1|}{|x_1^{(1)}|} \leqq 1.$$

---

since, if necessary, we can replace Σ by an infinite subsequence. If $\mu = 0$, then for all sufficiently large k

$$F(X^{(k)}) = F(X^{(k*)}),$$

so that the sequence $X^{(1)*}, X^{(2)*}, X^{(3)*}, \cdots$ has the same properties as Σ, contrary to the hypothesis on P_{n-1}.

Hence if

$$\frac{x_1^{(k+1)}}{x_1^{(k)}} = q^{(k)}, \qquad \text{then} \qquad \lim_{k \to \infty} | q^{(k)} | = 1,$$

so that for all sufficiently large k

$$| q^{(k)} | = 1.$$

Obviously

$$X^{*(k)} = X^{(k+1)} - q^{(k)} X^{(k)} = X^{(k+1)*} - q^{(k)} X^{(k)*}$$

lies in P_{n-1}, and for all large k

$$F(X^{(k)}) = F(q^{(k)} X^{(k)}) \neq F(X^{(k+1)}).$$

Hence

$$F(X^{*(k)}) = \max (F(X^{(k)}), F(X^{(k+1)})).$$

Therefore the sequence of positive numbers

$$F(X^{*(1)}), F(X^{*(2)}), F(X^{*(3)}), \cdots$$

contains an infinity of different elements and has the limit λ, so that again a contradiction is obtained.

Hence, if

$$X' = X - u_1 X^{(1)} = (0, x_2', \cdots, x_n'),$$

then

$$F(X') \leqq \max (F(X), |u_1| F(X^{(1)})) \leqq \tau,$$

and so X' also belongs to $C(\tau)$. Therefore

$$|x_2'| \leqq \xi_2,$$

so that

$$|u_2| = \frac{|x_2'|}{|x_2^{(2)}|} \leqq 1.$$

Continuing in this way, we obtain all inequalities

$$|u_1| \leqq 1, \cdots, |u_n| \leqq 1,$$

i.e. we have proved

$$\Phi_\tau(X) \leqq t, \qquad \text{if} \qquad F(X) \leqq \tau.$$

The domain defined by

$$\frac{1}{\tau} \Phi_\tau(X) = \max_{h=1,2,\cdots,n} \left(\left| \sum_{k=1}^{n} \alpha_{hk} x_k \right| \right) \leqq 1$$

is called a parallelepiped; our result may therefore be expressed in the form: *If the valuation $|x|$ is discrete, than every convex body $C(\tau)$ is a parallelepiped.*

As we have proved, the two domains

$$F(X) \leqq \tau \qquad \text{and} \qquad \Phi_\tau(X) \leqq \tau$$

are identical. In general, this does not imply the identity[5]

$$F(X) = \Phi_\tau(X)$$

for all X, and the function $\Phi_\tau(X)$ depends on τ. Suppose, however, that the set of values of $F(X)$ is the same as that of the values of $|x|$, and that τ is also an element of this set.[6] Then

$$\Phi_\tau(X) = \Phi(X)$$

becomes independent of τ, and for all X in P_n identically

$$F(X) = \Phi(X), \tag{10}$$

as follows easily from the property (B) of the distance functions.—

[5] E.g., if $\mathfrak{R} = \mathfrak{K}$ is the p-adic field ($p \geqq 3$), $n = 2$, and

$$F(X) = \max (|x_1|_p, 2|x_2|_p).$$

[6] It suffices to assume that $F(X)$ does not assume every positive value, and that the equation $F(X) = \tau$ has no solution.

Next assume that *the valuation* $|x|$ *is not discrete,* so that *its values lie everywhere dense on the positive real axis.* Now the n vectors

$$X_\epsilon^{(1)}, X_\epsilon^{(2)}, \cdots, X_\epsilon^{(n)}$$

will depend on ϵ, and so does the function

$$\Phi_{\tau\epsilon}(X) = \tau \max_{h=1,2,\cdots,n} (|u_{h\epsilon}|) = \tau \max_{h=1,2,\cdots,n} \left(\left|\sum_{k=1}^{n} \alpha_{hk\epsilon} x_k\right|\right).$$

Evidently

(11) $$F(X) \leqq \tau, \quad \text{if} \quad \Phi_{\tau\epsilon}(X) \leqq \tau.$$

Conversely, suppose $F(X) \leqq \tau$. Then

$$|x_1| \leqq \xi_1$$

and therefore

$$|u_{1\epsilon}| = \frac{|x_1|}{|x_{1\epsilon}^{(1)}|} < 1 + \epsilon.$$

Hence, if

$$X_\epsilon' = X - u_{1\epsilon} X_\epsilon^{(1)} = (0, x_{2\epsilon}', \cdots, x_{n\epsilon}'),$$

then

$$F(X_\epsilon') \leqq \max (F(X), |u_{1\epsilon}| F(X_\epsilon^{(1)})) < (1 + \epsilon)\tau.$$

There is a number α_ϵ in $\mathfrak{K}$ such that

$$F(X_\epsilon') \leqq |\alpha_\epsilon| \tau \leqq (1 + \epsilon)\tau, \quad \text{i.e.} \quad F(\alpha_\epsilon^{-1} X_\epsilon') \leqq \tau.$$

Hence

$$|\alpha_\epsilon^{-1} x_{2\epsilon}'| \leqq \xi_2, \qquad |x_{2\epsilon}'| \leqq (1 + \epsilon)\xi_2,$$

and therefore

$$|u_{2\epsilon}| = \frac{|x_{2\epsilon}'|}{|x_{2\epsilon}^{(2)}|} < (1 + \epsilon)^2,$$

so that, if

$$X_\epsilon'' = X_\epsilon' - u_{2\epsilon} X_\epsilon^{(2)} = X - (u_{1\epsilon} X_\epsilon^{(1)} + u_{2\epsilon} X_\epsilon^{(2)}) = (0, 0, x_{3\epsilon}'', \cdots, x_{n\epsilon}''),$$

then

$$F(X_\epsilon'') \leqq \max (F(X_\epsilon'), |u_{2\epsilon}| F(X_\epsilon^{(2)})) < (1 + \epsilon)^2\tau.$$

Continuing in the same way, we obtain the n inequalities

$$|u_{h\epsilon}| < (1 + \epsilon)^h \qquad (h = 1, 2, \cdots, n),$$

hence

(12) $$\Phi_{\tau\epsilon}(X) < (1 + \epsilon)^n\tau, \quad \text{if} \quad F(X) \leqq \tau.$$

From (11) and (12), since $\epsilon > 0$ is arbitrarily small:

If the valuation $|x|$ is everywhere dense on the positive axis, then the convex body $C(\tau)$ can be approximated arbitrarily near both from the inside and outside by means of parallelepipeds.

Take now, say $\tau = 1$ and put

$$\Phi_\epsilon(X) = \Phi_{1\epsilon}(X) = \max_{h=1,2,\cdots,n}\left(\left|\sum_{k=1}^{n} \alpha_{hk\epsilon} x_k\right|\right).$$

To every point X, there are two elements α and β of $\mathfrak{K}$, such that

$$\Phi_\epsilon(X) \leqq |\alpha| \leqq (1+\epsilon)\Phi_\epsilon(X) \qquad \text{and} \qquad F(X) \leqq |\beta| \leqq (1+\epsilon)F(X).$$

Hence from (11)

$$\Phi_\epsilon\left(\frac{X}{\alpha}\right) \leqq 1, \qquad F\left(\frac{X}{\alpha}\right) \leqq 1, \qquad F(X) \leqq |\alpha| \leqq (1+\epsilon)\Phi_\epsilon(X),$$

and from (12)

$$F\left(\frac{X}{\beta}\right) \leqq 1, \quad \Phi_\epsilon\left(\frac{X}{\beta}\right) \leqq (1+\epsilon)^n, \quad \Phi_\epsilon(X) \leqq (1+\epsilon)^n|\beta| \leqq (1+\epsilon)^{n+1}F(X),$$

and therefore uniformly in X

$$(1+\epsilon)^{-(n+1)}\Phi_\epsilon(X) \leqq F(X) \leqq (1+\epsilon)\Phi_\epsilon(X). \tag{13}$$

In general, these inequalities cannot be improved to an equation analogous to (10), e.g. if $F(X) = \tau$ has no solution.

4. The character of a convex set. If $F(X)$ is not special, then the set M of all solutions of $F(X) = 0$ contains elements other than $X = 0$. From (B) and (C), with X and Y also $aX + bY$ belongs to M, if a and b are elements of $\mathfrak{K}$. Hence M is a $\mathfrak{K}$-modul, say of dimension $n - m$. Obviously $m < n$; it is possible that $m = 0$, but then $F(X)$ vanishes identically and $C(\tau)$ is the whole space. Suppose therefore, that $1 \leqq m \leqq n - 1$, and let

$$P^{(m+1)}, P^{(m+2)}, \cdots, P^{(n)}$$

be $n - m\mathfrak{K}$-independent elements of M,

$$P^{(1)}, P^{(2)}, \cdots, P^{(m)}$$

m other points of P_n, so that the system of n vectors

$$P^{(1)}, P^{(2)}, \cdots, P^{(n)}$$

is still $\mathfrak{K}$-independent. Then every point X in P_n can be written as

$$X = v_1P^{(1)} + \cdots + v_nP^{(n)}$$

with elements $v_1, \cdots, v_n$ of $\mathfrak{K}$, viz.

$$v_h = \sum_{k=1}^{n} \beta_{hk} x_k \qquad (h = 1, 2, \cdots, n),$$

where the constant matrix in $\mathfrak{K}$

$$(\beta_{hk})_{h,k=1,2,\ldots,n}$$

has non-vanishing determinant. Since

$$F\left(\sum_{h=m+1}^{n} v_k P^{(h)}\right) = 0,$$

we have

$$F(X) = F\left(\sum_{h=1}^{m} v_h P^{(h)}\right) = \Psi(V),$$

where

$$\Psi(V) = \Psi(v_1, \cdots, v_m) = \Psi\left(\sum_{k=1}^{n} \beta_{1k} x_k, \cdots, \sum_{k=1}^{n} \beta_{nk} x_k\right)$$

is now obviously a special distance function in the m-dimensional space P_m of all points $V = (v_1, \cdots, v_m)$. Every convex set with $m > 0$ can therefore be considered as a cylinder, the basis of which is a convex body of $m < n$ dimensions.

5. The polar body of $C(\tau)$. Let $F(X)$ be the general distance function of §4, Y an arbitrary vector in P_n. Then we define a function $G(Y)$ by

(14) $G(0) = 0$; $G(Y) = \limsup (|XY|)$ *for all* X *with* $F(X) \leq 1$, *if* $Y \neq 0$.

In order to determine this function, let

$$Q^{(1)}, Q^{(2)}, \cdots, Q^{(n)}$$

be the n points in P_n, which satisfy the equations

$$P^{(h)} Q^{(k)} = \begin{cases} 1 & \text{for } h = k, \\ 0 & \text{for } h \neq k, \end{cases}$$

and write

$$Y = w_1 Q^{(1)} + \cdots + w_n Q^{(n)};$$

then

$$w_h = \sum_{k=1}^{n} \gamma_{hk} y_k \qquad (h = 1, 2, \cdots, n),$$

where the determinant of the matrix in $\mathfrak{K}$

$$(\gamma_{hk})_{h,k=1,2,\ldots,n}$$

does not vanish. Then

$$XY = v_1 w_1 + \cdots + v_n w_n .$$

Hence obviously

$$G(Y) = \infty, \qquad \text{unless} \qquad w_{m+1} = \cdots = w_n = 0.$$

Suppose therefore that

$$(15) \qquad w_{m+1} = w_{m+2} = \cdots = w_n = 0,$$

and put

$$G(Y) = \mathrm{X}(W),$$

where $W = (w_1, \cdots, w_m)$ is a vector in P_m. Then from (14),

$$(16) \quad \mathrm{X}(0) = 0; \quad \mathrm{X}(W) = \limsup\,(|\,VW\,|)\ \textit{for all } V \textit{ with } \Psi(V) \leqq 1, \quad \textit{if } W \neq 0,$$

so that the relation of $\mathrm{X}(W)$ to $\Psi(V)$ is the same as that of $G(Y)$ to $F(X)$. By §4, $\Psi(V)$ is a *special* distance function, and so is $\mathrm{X}(W)$, as follows easily from (16) and the properties (A′), (B), and (C) of $\Psi(V)$.

We call $G(Y)$ the *polar function to* $F(X)$; *for* $m < n$ *it is not itself a distance function*, but becomes one in the m-dimensional space (15), where it coincides with $\mathrm{X}(W)$. The set $C'(1/\tau)$: $G(Y) \leqq 1/\tau$, is further called the *polar set to* $C(\tau)$; it lies entirely in (15) and here is identical with the convex body $\mathrm{X}(W) = 1/\tau$.

Suppose now that $m = n$, i.e. both $F(X)$ and $G(Y)$ are special distance functions; then the polar set $C'(1/\tau)$ becomes a convex body. We shall prove that in this case *the relation between* $F(X)$ *and* $G(Y)$ *is reciprocal*, i.e. $F(X)$ is the polar function to $G(Y)$ and $C(\tau)$ the polar body to $C'(1/\tau)$.

This assertion is evident, if $F(X) = |\,X\,|$, for then obviously $G(Y) = |Y|$. Further let

$$\Omega = (a_{hk})_{h,k=1,2,\ldots,n}\,; \qquad \Omega^K = (a_{hk}^K)_{h,k=1,2,\ldots,n}$$

be an arbitrary matrix in $\mathfrak{K}$ with nonvanishing determinant, and its complementary matrix, so that for all X and Y the scalar product[7]

$$\Omega X \cdot \Omega^K Y = XY.$$

Then the transformed distance functions $G'(Y) = G(\Omega^K Y)$ and $F'(X) = F(\Omega X)$ have still the property that the first one is polar to the second, since

$$\begin{aligned} G'(Y) = G(\Omega^K Y) &= \limsup_{F(X)\leqq 1}\,(|\,X\cdot\Omega^K Y\,|) \\ &= \limsup_{F(\Omega X)\leqq 1}\,(|\,\Omega X\cdot\Omega^K Y\,|) = \limsup_{F'(X)\leqq 1}\,(|\,XY\,|). \end{aligned}$$

Further, if $F_1(X)$ and $F_2(X)$ are two distance functions such that for all X

$$F_1(X) \leqq F_2(X),$$

[7] The vector $X' = (x_1', \cdots, x_n') = \Omega X$ is defined by $x_h' = \sum_{k=1}^{n} a_{hk}x_k$ for $h = 1, 2, \cdots, n$.

then the polar distance functions $G_1(Y)$ and $G_2(Y)$ satisfy the inverted inequality

$$G_1(Y) \geqq G_2(Y).$$

We distinguish now the same two cases as in §3. If the valuation $|x|$ is discrete, then we showed the existence of a matrix

$$A = (\alpha_{hk})_{h,k=1,2,\dots,n}$$

in $\mathfrak{K}$ with determinant different from zero, such that

$$F(X) = \Phi(X) = |AX|$$

identically in X. The polar function to $F(X)$ is therefore

$$G(Y) = |A^K Y|,$$

and since $(A^K)^K = A$, the statement follows at once.—In this case, the definition of $G(Y)$ can obviously be replaced by the simpler one:

$$G(Y) = \max_{X \neq 0} \frac{|XY|}{F(X)}. \tag{17}$$

Secondly, let $|x|$ be everywhere dense on the positive real axis. Then to every $\delta > 0$, there are two matrices

$$A_1 = (\alpha_{hk}^{(1)})_{h,k=1,2,\dots,n} \quad \text{and} \quad A_2 = (\alpha_{hk}^{(2)})_{h,k=1,2,\dots,n}$$

in $\mathfrak{K}$ with non-vanishing determinants, such that if

$$F_1(X) = |A_1X|, \qquad F_2(X) = |A_2X|,$$

then for all X

$$F_1(X) \leqq F(X) \leqq F_2(X) \leqq (1 + \delta)F_1(X),$$

as follows easily from (13). Hence if

$$G_1(Y) = |A_1^K Y|, \qquad G_2(Y) = |A_2^K Y|$$

are the polar functions to $F(X)$ and $F(X)$, then also

$$G_2(Y) \leqq G(Y) \leqq G_1(Y);$$

and[8]

$$G_2(Y) \leqq (1 + 2\delta)G_1(Y),$$

[8] There is a number α in $\mathfrak{K}$ such that

$$1 + \delta \leqq |\alpha| \leqq 1 + 2\delta.$$

Then by hypothesis

$$F_1(X) \leqq (1 + \delta)F_2(X) \leqq F_2(\alpha X).$$

Hence

$$\frac{1}{1 + 2\delta} G_2(Y) \leqq G_2\left(\frac{Y}{\alpha}\right) \leqq G_1(Y),$$

since the polar function to $F_2(\alpha X)$ is $G_2\left(\frac{Y}{\alpha}\right)$.

for all Y. Since δ can be taken arbitrarily small, the assertion follows again for the same reason.—In this case, the definition of $G(Y)$ is easily replaced by

$$G(Y) = \limsup_{X \neq 0} \frac{|XY|}{F(X)}. \tag{17'}$$

By the proved reciprocity of $F(X)$ and $G(Y)$, the formulae (17) and (17′) remain true if $G(Y)$ is replaced by $F(X)$ and vice versa.

II. "Geometry of Numbers" in a Domain of Power Series

6. Notation. We specialize now the fields $\mathfrak{R}$ and $\mathfrak{K}$ of §1, and denote by

- $\mathfrak{k}$ *an arbitrary field,*
- z *an indeterminate,*
- $\mathfrak{T} = k[z]$ *the ring of all polynomials in* z *with coefficients in* $\mathfrak{k}$,
- $\mathfrak{R} = k(z)$ *the quotient field of* $\mathfrak{T}$, i.e. *the field of all rational functions in* z *with coefficients in* $\mathfrak{k}$,
- $|x|$ *the special valuation of* $\mathfrak{R}$ *defined by*

$$|x| = \begin{cases} 0, & \text{if } x = 0, \\ e^f, & \text{if } x \neq 0 \text{ is of order } f,^{9} \end{cases}$$

- $\mathfrak{K}$ *the perfect extension of* $\mathfrak{R}$ *with respect to this valuation, i.e. the field of all formal Laurent series*

$$x = \alpha_f z^f + \alpha_{f-1} z^{f-1} + \alpha_{f-2} z^{f-2} + \cdots$$

with coefficients in $\mathfrak{k}$; *if* α_f *is the non-vanishing coefficient with highest index* $\gtreqless 0$, *then* $|x| = e^f$,

- Λ_n *the set of all "lattice points" in* P_n, *i.e. that of all points with coordinates in* $\mathfrak{T}$.

The valuation $|x|$ is by definition a power of e with integral exponent. We assume the same for all distance functions which we consider from now onwards, and we shall consider only convex sets or bodies $C(\tau)$, where τ is an exact power of e, say $\tau = e^t$.

7. The volume V of a convex body $C(1)$. Let $F(X)$ be a special distance function, $C(e^t)$ the convex body $F(X) \leqq e^t$, where t is an arbitrary integer. It is obvious that the set $\mathfrak{m}(t)$ of all lattice points in $C(e^t)$ forms a $\mathfrak{k}$-modul. In the special case $F(X) = |X|$, this set has exactly

$$M_0(t) = n(t + 1)$$

$\mathfrak{k}$-independent elements. Hence, by the inequalities (5) and (6), $\mathfrak{m}(t)$ has always a finite dimension $M(t)$, and this dimension is certainly positive for large t.

[9] The order of a rational function is the degree of its numerator minus the degree of its denominator.

Obviously

$$M_0(t + 1) = M_0(t) + n. \tag{18}$$

Suppose that t is already so large that

$$e^{t+1} \geqq \Gamma.$$

Then a lattice point in $C(e^{t+1})$ can be written as

$$X = X_0 + zX_1,$$

where X_0 and X_1 are again lattice points, and the coordinates of X_0 lie in $\mathfrak{k}$, i.e.

$$|X_0| \leqq 1, \qquad F(X_0) \leqq \Gamma \leqq e^{t+1}.$$

Hence

$$F(zX_1) \leqq \max(F(X), F(X_0)) \leqq e^{t+1}, \qquad F(X_1) \leqq e^t,$$

so that X_1 lies in $\mathfrak{m}(t)$. Conversely, if X_1 belongs to $\mathfrak{m}(t)$, then

$$F(X) \leqq \max(F(zX_1), F(X_0)) \leqq e^{t+1}.$$

Now the two vectors X_0 and zX_1, where X_0 and X_1 are lattice points and $|x_0| \leqq 1$, are $\mathfrak{k}$-independent, and the X_0 form a $\mathfrak{k}$-modul of dimension n. Hence

$$M(t + 1) = M(t) + n. \tag{19}$$

The two equations (18) and (19) show that for large t, the function $M(t) - M_0(t)$ of t is independent of t. Hence the limit

$$V = \lim_{t\to\infty} e^{M(t)-M_0(t)} \tag{20}$$

exists; it is called the *volume of the convex body* $C(1)$.[10] In particular, if $F(X) = |X|$, then obviously $V = 1$.

8. The invariance of V. Let

$$\Omega = (a_{hk})_{h,k=1,2,\ldots,n} \quad \text{and} \quad \Omega^I = (a^I_{hk})_{h,k=1,2,\ldots,n}$$

be a matrix with elements in $\mathfrak{K}$ and determinant $D \neq 0$, and its inverse matrix. The linear transformation

$$Y = \Omega X \quad \text{or} \quad X = \Omega^I Y$$

changes $F(X)$ into the new distance function

$$F'(Y) = F(X) = F(\Omega^I Y);$$

let $C'(e^t)$ be the corresponding convex body $F'(Y) \leqq e^t$, and V' the volume of $C'(1)$. *Then*

$$V' = |D| V. \tag{21}$$

[10] This definition is analogous to that of the volume of a body by means of lattice points in an ordinary real space.

PROOF: We denote by $\mathfrak{m}'(t)$ the $\mathfrak{k}$-modul of all lattice points in $C'(e^t)$, by $M'(t)$ the dimension of $\mathfrak{m}'(t)$, and prove the statement in a number of steps.

1: The elements of Ω lie in $\mathfrak{T}$, and D belongs to $\mathfrak{k}$.

The formulae $Y = \Omega X$, $X = \Omega^I Y$ establish a (1, 1)-correspondence between the elements X of $\mathfrak{m}(t)$ and Y of $\mathfrak{m}'(t)$. Obviously, this correspondence changes every linear relation

$$\alpha_1 X^{(1)} + \cdots + \alpha_r X^{(r)} = 0$$

with coefficients in $\mathfrak{k}$ into the identical relation in the Y's, and vice versa; therefore $\mathfrak{k}$-independent elements of $\mathfrak{m}(t)$ or $\mathfrak{m}'(t)$ are transformed into $\mathfrak{k}$-independent members of the other modul. Hence both moduls have the same dimension: $M(t) = M'(t)$, q.e.d.

2: Ω is a triangle matrix

$$\Omega = \begin{pmatrix} a_{11} & & & \\ a_{21} & a_{22} & & 0 \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

with elements in $\mathfrak{T}$ and determinant

$$D = a_{11}a_{22} \cdots a_{nn} \neq 0.$$

The equation $Y = \Omega X$ denotes that

$$\begin{aligned} y_1 &= a_{11}x_1, \\ y_2 &= a_{21}x_1 + a_{22}x_2, \\ &\cdots\cdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n; \end{aligned}$$

hence every lattice point Y can be written as[11]

$$Y = \Omega X^* + Y^*,$$

where X^* and Y^* are again lattice points and $Y^* = (y_1^*, \cdots, y_n^*)$ satisfies the inequalities

$$|y_1^*| < |a_{11}|, |y_2^*| < |a_{22}|, \cdots, |y_n^*| < |a_{nn}|.$$

Therefore

$$|Y^*| \leqq c_1, \text{ i.e. } F'(Y^*) \leqq c_1\Gamma',$$

where c_1 is a positive constant depending only on Ω, and Γ' is the constant in (5) belonging to $F'(Y)$. The set of all vectors Y^* forms a $\mathfrak{k}$-modul $\mathfrak{m}^*$ of dimension d, where

$$e^d = |a_{11}||a_{22}| \cdots |a_{nn}| = |D|.$$

---

[11] We use the trivial lemma: "*To a and $b = 0$ in $\mathfrak{T}$ there is a q and an r in $\mathfrak{T}$, such that $a = bq + r$ and $|r| < |b|$.*

Let t be so large that

$$e^t \geqq c_1\Gamma'.$$

Then for X^* in $\mathfrak{m}(t)$

$$F'(Y) = F(\Omega^I Y) = F(X^* + \Omega^I Y^*) \leqq \max(F(X^*), F'(Y^*)) \leqq e^t,$$

and conversely for Y in $\mathfrak{m}'(t)$

$$F(X^* + \Omega^I Y^*) \leqq e^t, \text{ i.e. } F(X^*) \leqq \max(F(X^* + \Omega^I Y^*), F'(Y^*)) \leqq e^t.$$

There is therefore a (1, 1)-correspondence between the elements Y of $\mathfrak{m}'(t)$ and the pairs (X^*, Y^*) of one element X^* of $\mathfrak{m}(t)$ and one element Y^* of $\mathfrak{m}^*$. Hence $M'(t) = M(t) + d$, q.e.d.

3: The elements of Ω belong to $\mathfrak{T}$.

The result follows immediately from the two previous steps, since Ω, as is well known,[12] can be written as $\Omega = \Omega_1\Omega_2$, where the two factors are of the classes 1 and 2.

4: The elements of Ω lie in $\mathfrak{R}$.

Now $\Omega = \Omega_a\Omega_b^I$, where both Ω_a and Ω_b are of the class 3, so that the statement follows at once.

5: Ω has elements in $\mathfrak{K}$, such that

$$|D| = 1, \qquad |a_{hk}| \leqq 1 \qquad (h, k = 1, 2, \cdots, n).$$

Then the same inequalities hold for the inverse matrix Ω^I, so that for every point X

$$|\Omega X| \leqq |X|, \qquad |X| = |\Omega^I\Omega X| \leqq |\Omega X|,$$

and therefore

$$|X| = |\Omega X| = |\Omega^I X|.$$

Now to every lattice point X there is a second lattice point Y such that with a suitable point Y^*

$$\Omega X = Y + Y^*, \qquad |Y^*| < 1;$$

then conversely

$$\Omega^I Y = X + X^*, \qquad |X^*| < 1,$$

and

$$X^* = -\Omega^I Y^*, \qquad \Omega X^* = -Y^*.$$

The relation between X and Y is therefore a (1, 1)-correspondence which obviously leaves invariant the property of $\mathfrak{k}$-independence. Suppose that

$$e^t \geqq \Gamma.$$

[12] This can be proved, e.g. by a method analogous to Minkowki's "adaptation" of a lattice; Geometrie der Zahlen §46.

Then for X in $\mathfrak{m}(t)$

$$F(X^*) < \Gamma \leqq e^t,$$

and therefore

$$F'(Y) = F(\Omega^I Y) = F(X + X^*) \leqq \max(F(X), F(X^*)) \leqq e^t,$$

so that Y lies in $\mathfrak{m}'(t)$; conversely, if Y belongs to $\mathfrak{m}'(t)$. then X is an element of $\mathfrak{m}(t)$. Hence $M(t) = M'(t)$, q.e.d.

6: Finally, let Ω have elements in $\mathfrak{K}$. Then it can be split into

$$\Omega = \Omega_4 + \Omega^*$$

where Ω_4 is of the class 4, while the elements of Ω^* lie in $\mathfrak{K}$ and have so small values that

$$\Omega_5 = \Omega_4^I \Omega$$

is of the class 5. Then the result follows at once, since $\Omega = \Omega_4 \Omega_5$.

Two conclusions are immediate from (21). The convex body $C(e^t)$, i.e. $F(z^{-t}X') \leqq 1$, is obtained from $C(1)$ by the transformation $X' = z^t X$; hence it has the volume $V(e^t) = e^{nt}V$. Secondly, *let $G(Y)$ be the polar distance function to $F(X)$, and V' the volume of the convex body $C'(1)$*, i.e. $G(Y) \leqq 1$. *Then V and V' are related by the equation*

$$VV' = 1. \tag{22}$$

For by §5, there is a matrix A with non-vanishing determinant, such that

$$F(X) = |AX| \quad \text{and} \quad G(Y) = |A^K Y|,$$

hence

$$V = (|A|)^{-1} \quad \text{and} \quad V' = (|A^k|)^{-1} = |A|;$$

the statement is therefore obvious.

9. The minima of $F(X)$. To the distance function $F(X)$, there exist n $\mathfrak{K}$-independent lattice points

$$X^{(k)} = (x_1^{(k)}, \cdots, x_n^{(k)}) \qquad (k = 1, 2, \cdots, n),$$

such that

$F(X^{(1)}) = \sigma^{(1)} = e^{g_1}$ *is the minimum of $F(X)$ in all lattice points $X \neq 0$,*

$F(X^{(2)}) = \sigma^{(2)} = e^{g_2}$ *is the minimum of $F(X)$ in all lattice points X which are $\mathfrak{K}$-independent of $X^{(1)}$*, etc., and finally

$F(X^{(n)}) = \sigma^{(n)} = e^{g_n}$ *is the minimum of $F(X)$ in all lattice points X which are $\mathfrak{K}$-independent of $X^{(1)}, X^{(2)}, \cdots, X^{(n-1)}$.*

The numbers $\sigma^{(1)}, \sigma^{(2)}, \cdots, \sigma^{(n)}$ are called *the n successive minima of $F(X)$*. By this construction, the determinant

$$D = |x_h^{(k)}|_{h,k=1,2,\cdots,n}$$

lies in $\mathfrak{T}$ and does not vanish; further obviously

$$(23) \qquad 0 < \sigma^{(1)} \leqq \sigma^{(2)} \leqq \cdots \leqq \sigma^{(n)} \quad \text{and} \quad g_1 \leqq g_2 \leqq \cdots \leqq g_n .$$

We shall prove the two equations

$$(24) \qquad |D| = 1,$$

$$(25) \qquad \sigma^{(1)}\sigma^{(2)} \cdots \sigma^{(n)} = \frac{1}{V};$$

in the second one, V is again the volume of $C(1)$. Thus, in particular, *D is an element of $\mathfrak{k}$, and may obviously be taken as equal to* 1.

A: PROOF OF (24). Every point X in P_n can be written as

$$X = y_1X^{(1)} + \cdots + y_nX^{(n)},$$

where the y's are elements of $\mathfrak{K}$. Then the coordinates x_h of X are linear functions with determinant D of the coordinates y_h of $Y = (y_1, \cdots, y_n)$. We define a new distance function $\Pi(X)$ by

$$\Pi(X) = |Y|.$$

By (21), the convex body $\Pi(X) \leqq 1$ has the volume $|D|$; we determine it in the following way:

If X is a lattice point, then Y also has its coordinates y_h in $\mathfrak{T}$. For since with Y also X is obviously a lattice point, we may assume without loss of generality that

$$(26) \qquad \Pi(X) = |Y| < 1,$$

and have to show that no lattice point $X \neq 0$ satisfies this inequality. Let m, where $1 \leqq m \leqq n$, be the greatest index for which $y_m \neq 0$. Then

$$X = \sum_{h=1}^{m} y_h X^{(h)}, \qquad X^{(1)}, \cdots, X^{(m-1)}$$

are $\mathfrak{R}$-independent lattice points, and by (26)

$$F(X) \leqq \max(|y_1| F(X^{(1)}), \cdots, |y_m| F(X^{(m)})) < \sigma^{(m)},$$

in contradiction to the minimum property of $\sigma^{(m)}$.

Hence there are exactly $M_1(t) = n(t+1)$ $\mathfrak{k}$-independent lattice points such that $\Pi(X) \leqq e^t$, viz. all points corresponding to a basis of $\mathfrak{k}$-independent points Y with $|Y| \leqq e^t$. Therefore

$$|D| = \lim_{t\to\infty} e^{M_1(t)-M_0(t)} = 1, \qquad \text{q.e.d.}$$

B: PROOF OF (25). Now we use the fact that every point X in P_n can be written as

$$X = y_1 z^{-g_1} X^{(1)} + \cdots + y_n z^{-g_n} X^{(n)},$$

where the y's belong to $\mathfrak{K}$. Let $\Sigma(X)$ be the distance function given by

$$\Sigma(X) = | Y |.$$

Since

$$F(z^{-g_h} X^{(h)}) = 1 \qquad (h = 1, 2, \cdots, n),$$

obviously

$$F(X) \leqq 1, \quad \text{if} \quad \Sigma(X) \leqq 1.$$

But the converse is also true: If

$$F(X) \leqq 1, \quad \text{then} \quad \Sigma(X) \leqq 1,$$

and therefore evidently

$$F(X) = \Sigma(X) = | Y |,$$

identically in X.

For suppose that on the contrary for a certain point X in P_n

$$F(X) \leqq 1, \quad \text{but} \quad \Sigma(X) > 1.$$

Then let m with $1 \leqq m \leqq n$ be the greatest index for which $| y_m | > 1$; hence if $m < n$

$$| y_{m+1} | \leqq 1, \cdots, | y_n | \leqq 1.$$

Write

$$y_h = z y_h^* + y_h^{**} \qquad (h = 1, 2, \cdots, n),$$

where the y_h^* are elements of $\mathfrak{T}$, the y_h^{**} elements of $\mathfrak{K}$, and

$$y_m^* \neq 0, \qquad y_{m+1}^* = \cdots = y_n^* = 0, \qquad | y_1^{**} | \leqq 1, \cdots, | y_n^{**} | \leqq 1,$$

and put

$$Y^* = (y_1^*, \cdots, y_n^*), \qquad Y^{**} = (y_1^{**}, \cdots, y_n^{**}),$$

so that

$$Y = zY^* + Y^{**}.$$

Obviously, Y^* is a lattice point, Y^{**} a point such that $| Y^{**} | \leqq 1$. Also write

$$X^* = \sum_{h=1}^{n} y_h^* z^{-g_h} X^{(h)} = \sum_{h=1}^{m} y_h^* z^{-g_h} X^{(h)}, \qquad X^{**} = \sum_{h=1}^{n} y_h^{**} z^{-g_h} X^{(h)},$$

so that

$$X = zX^* + X^{**}.$$

Then from $\Sigma(X^{**}) = | Y^{**} | \leqq 1$,

$$F(X^{**}) \leqq 1.$$

Hence

$$F(zX^*) \leqq \max (F(X), F(X^{**})) \leqq 1, \qquad F(X^*) < 1,$$

and

$$F(X^0) < \sigma^{(m)},$$

where $X^0 = z^{g_m}X^*$. This inequality, however, is impossible, since the m lattice points

$$X^0 = \sum_{h=1}^{m} y_h^* z^{g_m - g_h} X^{(h)}, \qquad X^{(1)}, \cdots, X^{(m-1)}$$

are $\mathfrak{K}$-independent, so that by the minimum property of $\sigma^{(m)}$

$$F(X^{(0)}) \geqq \sigma^{(m)}.$$

Therefore (27) is true, so that by the invariance theorem of §8

$$V = \frac{|D|}{\sigma^{(1)}\sigma^{(2)} \cdots \sigma^{(n)}} = (\sigma^{(1)}\sigma^{(2)} \cdots \sigma^{(n)})^{-1},$$

since the transformation of X into Y has the determinant

$$Dz^{-(g_1+g_2+\cdots+g_n)}.$$

The equation (25) is therefore proved.

From this equation and from (23) in particular

$$\sigma^{(1)} \leqq V^{-1/n};$$

i.e. *to every distance function $F(X)$ there is a lattice point $X \neq 0$ such that*

$$F(X) \leqq \frac{1}{\sqrt[n]{V}}.$$

Here equality holds if and only if all minima

$$\sigma^{(1)} = \sigma^{(2)} = \cdots = \sigma^{(n)},$$

thus certainly not, if V is not an integral power of e^n.

10. The relations between the minima of $F(X)$ and $G(Y)$. To the n lattice points $X^{(1)}, X^{(2)}, \cdots, X^{(n)}$ defined in the last paragraph, we construct n points $Y^{(1)}, Y^{(2)}, \cdots, Y^{(n)}$ satisfying

$$(27) \qquad X^{(h)}Y^{(n-k+1)} = \begin{cases} 1 & \text{for } h = k, \\ 0 & \text{for } h \neq k; \end{cases}$$

since $|D| = 1$, these points are lattice points. We further define n positive numbers

$$(28) \qquad \tau^{(h)} = \frac{1}{\sigma^{(n-h+1)}} = e^{j_h} \qquad (h = 1, 2, \cdots, n),$$

so that

(29) $$0 < \tau^{(1)} \leqq \tau^{(2)} \leqq \cdots \leqq \tau^{(n)} \quad \text{and} \quad j_1 \leqq j_2 \leqq \cdots \leqq j_n .$$

Then $F(X)$ and the polar function $G(Y)$ can be written as

(30) $$F(X) = \max_{h=1,2,\cdots,n} (\sigma^{(h)} \mid XY^{(n-h+1)} \mid),$$

(31) $$G(Y) = \max_{h=1,2,\cdots,n} (\tau^{(h)} \mid YX^{(n-h+1)} \mid),$$

thus in an entirely symmetrical way. For we proved in the preceding paragraph that if X is written as

(32) $$X = \sum_{h=1}^{n} y_h z^{-g_h} X^{(h)},$$

then

$$F(X) = \mid Y \mid, \qquad Y = (y_1, y_2, \cdots, y_n).$$

But by multiplying (32) scalar with $Y^{(n)}, \cdots, Y^{(1)}$, we get by (27)

$$y_h = z^{g_h} \cdot (XY^{(n-h+1)}) \qquad (h = 1, 2, \cdots, n)$$

and therefore (30). The formula (31) is a consequence of (30) by the results in §5.[13]

From (27) and (31)

(33) $$G(Y^{(h)}) = \tau^{(h)} = e^{j_h}.$$

We prove now that these numbers $\tau^{(h)}$ in their natural order are the n successive minima of $G(Y)$ in Λ_n. Obviously it suffices to show that *if*

$$Z^{(1)}, Z^{(2)}, \cdots, Z^{(n)}.$$

are any n $\mathfrak{K}$-independent lattice points, such that

$$G(Z^{(1)}) \leqq G(Z^{(2)}) \leqq \cdots \leqq G(Z^{(n)}),$$

---

[13] We can prove (31) directly in the following way: Obviously

$$X = \sum_{h=1}^{n} (XY^{(n-h+1)})X^{(h)},$$

where the brackets are again the scalar products. Hence from (14)

$$G(Y) = \max (\mid XY \mid) = \max \left(\left| \sum_{h=1}^{n} (XY^{(n-h+1)})(X^{(h)}Y) \right| \right),$$

where the maximum extends over all points X of $C(1)$, i.e. for which

$$\mid XY^{(n-h+1)} \mid \leqq \frac{1}{\sigma^{(h)}} = \tau^{(n-h+1)} \qquad (h = 1, 2, \cdots, n).$$

By choosing X such that there is equality in one of these conditions, but that all other scalar products $XY^{(n-h+1)}$ vanish, the assertion follows after replacing h by $n - h + 1$.

then[14]

$$G(Z^{(h)} \geqq G(Y^{(h)}) = \tau^{(h)}.$$

Consider the $n + 1$ vectors

$$X^{(1)}, X^{(2)}, \cdots, X^{(n-h+1)}, \qquad Z^{(1)}, Z^{(2)}, \cdots, Z^{(h)}.$$

At most n of these are $\mathfrak{K}$-independent; hence the scalar products

$$X^{(i)} Z^{(j)} \qquad \begin{pmatrix} i = 1, 2, \cdots, n - h + 1 \\ j = 1, 2, \cdots, \quad h \end{pmatrix}$$

do not all vanish simultaneously, and at least one of them, say $X^{(i)}Z^{(j)}$, is different from zero. Since it is an element of $\mathfrak{T}$, therefore

$$| X^{(i)}Z^{(j)} | \geqq 1.$$

Now by (17)

$$| XY | \leqq F(X)G(Y),$$

for all points X and Y. Therefore

$$1 \leqq | X^{(i)} Z^{(j)} | \leqq F(X^{(i)})G(Z^{(j)}) \leqq F(X^{(n-h+1)})G(Z^{(h)}) = \frac{1}{\tau^{(h)}} G(Z^{(h)}),$$

as was to be proved.

From (28) and (29) in particular

$$\sigma^{(1)} \leqq \left(\frac{\tau^{(1)}}{V}\right)^{1/n-1} \quad \text{and} \quad \tau^{(1)} \leqq (\sigma^{(1)}V)^{1/n-1}, \tag{34}$$

so that if the minimum of $F(X)$ in $\mathfrak{T}$ is small, then the same is true for that of $G(Y)$, and vice versa.

11. The relation between the homogeneous and the inhomogeneous problem. The reciprocity formulae of the preceding paragraph can be applied to inhomogeneous problems. Let P be an arbitrary point in P_n which is not necessarily a lattice point; it can be written as

$$P = p_1X^{(1)} + \cdots + p_nX^{(n)}$$

where the p's lie in $\mathfrak{K}$. Put

$$p_h = -x_h + r_h \qquad (h = 1, 2, \cdots, n),$$

where x_h is an element of $\mathfrak{T}$ and

$$|r_h| \leqq \frac{1}{e} \qquad (h = 1, 2, \cdots, n).$$

[14] The minima $\sigma^{(h)}$ of $F(X)$ have the analogous property.

Then the lattice point $X = (x_1, \cdots, x_n)$ satisfies the inequality

$$F(P + X) = F\left(\sum_{h=1}^{n} r_h X^{(h)}\right) \leqq \frac{\sigma^{(n)}}{e},$$

or by (28)

$$F(P + X) \leqq \frac{1}{e\tau^{(1)}}. \tag{35}$$

This inequality cannot in general be improved, since

$$F\left(\frac{1}{z} X^{(n)} + X\right) \geqq \frac{1}{e\tau^{(1)}} \tag{36}$$

for all lattice points X, as follows immediately from the $\mathfrak{K}$-independence of the n vectors

$$X^{(1)}, X^{(2)}, \cdots, X^{(n-1)}, X^{(n)} + zX.$$

These two inequalities (35) *and* (36) *relate the inhomogeneous F-problem to the homogeneous G-problem,* in analogy with similar relations in many parts of mathematics.

As an application, consider the two polar distance functions

$$F(X) = \max\left(|\alpha_1 x_n - x_1|, \cdots, |\alpha_{n-1} x_n - x_{n-1}|, e^{-t} |x_n|\right),$$

$$G(Y) = \max\left(|y_1|, \cdots, |y_{n-1}|, e^{t} |\alpha_1 y_1 + \cdots + \alpha_{n-1} y_{n-1} + y_n|\right),$$

where t is a positive integer. Assume that the numbers $1, \alpha_1, \cdots, \alpha_{n-1}$ are $\mathfrak{R}$-independent, so that for all lattice points $Y = (y_1, \cdots, y_n) \neq 0$

$$\alpha_1 y_1 + \cdots + \alpha_{n-1} y_{n-1} + y_n \neq 0.$$

Then, as $t \to \infty$, the first minimum $\tau^{(1)}$ of $G(Y)$

$$\tau^{(1)} \to \infty.$$

Hence by (35), *for every* $\epsilon > 0$ *and for every point* $P = (p_1, \cdots, p_n)$ *there is a lattice point* $X = (x_1, \cdots, x_n)$ *satisfying the inequalities*

$$|\alpha_1 x_n - x_1 + p_1| < \epsilon, \cdots, |\alpha_{n-1} x_n - x_{n-1} + p_{n-1}| < \epsilon.$$

Thus we have established a result analogous to *Kronecker's* theorem.

12. A property of matrices. *Let*

$$\Omega = (a_{hk})_{h,k=1,2,\cdots,n}$$

be a matrix in $\mathfrak{K}$ *with determinant* 1; *then there is a matrix*

$$U = (u_{hk})_{h,k=1,2,\cdots,n}$$

with elements in $\mathfrak{T}$ *and determinant* 1, *such that the product matrix*

$$\Omega U = \Omega^* = (a^*_{hk})_{h,k=1,2,\cdots,n}$$

satisfies the equation

$$\prod_{h=1}^{n} \max_{k=1,2,\cdots,n} (|\, a^*_{hk} \,|) = 1.$$

PROOF:[15] To the convex body $C(1)$ belonging to the distance function

$$F(X) = \max_{h=1,2,\cdots,n} \left(\left| \sum_{k=1}^{n} a_{hk} x_k \right| \right),$$

there are n lattice points $X^{(1)}, X^{(2)}, \cdots, X^{(n)}$ of determinant $D = 1$, such that the n minima

$$F(X^{(h)}) = \sigma^{(h)} \qquad (h = 1, 2, \cdots, n)$$

satisfy

$$0 < \sigma^{(1)} \leqq \sigma^{(2)} \leqq \cdots \leqq \sigma^{(n)}, \qquad \sigma^{(1)}\sigma^{(2)} \cdots \sigma^{(n)} = 1.$$

Let $X^{(k)} = (x_1^{(k)}, \cdots, x_n^{(k)})$, and X be the matrix

$$\mathrm{X} = (x_h^{(k)})_{h,k=1,2,\cdots,n}$$

with elements in $\mathfrak{T}$ and determinant 1. We introduce new coordinates $y_1, \cdots, y_n$ by putting

$$X = y_1 X^{(1)} + \cdots + y_n X^{(n)}, \text{ i.e., } x_h = \sum_{k=1}^{n} x_h^{(k)} x_k \qquad (h = 1, 2, \cdots, n);$$

then $F(X)$ changes into

$$F(X) = F'(Y) = \max_{h=1,2,\cdots,n} \left(\left| \sum_{k=1}^{n} a'_{hk} y_k \right| \right),$$

where

$$\Omega' = (a'_{hk})_{h,k=1,2,\cdots,n} = \Omega \mathrm{X}.$$

The n points $X = X^{(h)}$ are transformed into $Y = E^{(h)}$ $(h = 1, 2, \cdots, n)$; hence

$$F'(E^{(h)}) = \sigma^{(h)} \qquad (h = 1, 2, \cdots, n),$$

that is

$$\max_{h=1,2,\cdots,n} (|\, a'_{hk} \,|) = \sigma^{(k)} \qquad (k = 1, 2, \cdots, n). \tag{37}$$

[15] An analogous theorem in the real field was proved some time ago by C. L. Siegel in a letter to L. J. Mordell. The present proof and theorem, though not stated in Siegel's paper, are obtained from it with only slight changes by making use of the results in §9.

Hence every minor Δ_m of order m formed from the m first columns and m arbitrary rows of Ω' satisfies the inequality

$$|\Delta| \leqq \sigma^{(1)}\sigma^{(2)} \cdots \sigma^{(m)}. \tag{38}$$

On the other hand, any determinant Δ of order m can be written as

$$\Delta = \sum_{h=1}^{n} a_h \delta_h,$$

where the a_h are the elements of its last column, and the δ_h their cofactors; therefore

$$\max_{h=1,2,\cdots,n} (|\delta_h|) \geqq |\Delta| \{\max_{h=1,2,\cdots,n} (|a_h|)\}^{-1}.$$

We apply this inequality repeatedly to the determinant

$$\Delta_n = 1 = \sigma^{(1)}\sigma^{(2)} \cdots \sigma^{(n)}$$

of Ω' and use (37) and (38); then it follows that *there exists*

an $(n-1)^{\text{th}}$ *order minor* Δ_{n-1} *of* Δ_n *formed from the* $n-1$ *first columns of* Ω' *and satisfying*

$$|\Delta_{n-1}| = \sigma^{(1)}\sigma^{(2)} \cdots \sigma^{(n-1)};$$

an $(n-2)^{\text{th}}$ *order minor* Δ_{n-2} *of* Δ_{n-1} *formed from the* $n-2$ *first columns of* Ω' *and satisfying*

$$|\Delta_{n-2}| = \sigma^{(1)}\sigma^{(2)} \cdots \sigma^{(n-2)};$$

etc.; *a second order minor* Δ_2 *of* Δ_3 *formed from the two first columns of* Ω' *and satisfying*

$$|\Delta_2| = \sigma^{(1)}\sigma^{(2)};$$

and finally an element Δ_1 *of* Δ_2 *lying in the first column of* Ω' *and satisfying*

$$|\Delta_1| = \sigma^{(1)}.$$

Without loss of generality, we may assume that the determinants so constructed are exactly the principle determinants

$$\Delta_r = |a'_{hk}|_{h,k=1,2,\cdots,r} \qquad (r = 1, 2, \cdots, n).$$

We shall now construct a set of matrices of order n

$$U_m = \begin{pmatrix} 1 & 0 & \cdots & 0 & g_1^{(m)} & 0 & \cdots & 0 \\ & 1 & \cdots & 0 & g_2^{(m)} & 0 & \cdots & 0 \\ & & \ddots & \vdots & \vdots & \vdots & & \vdots \\ & & & 1 & g_{m-1}^{(m)} & 0 & \cdots & 0 \\ & & & & 1 & 0 & \cdots & 0 \\ & O & & & & 1 & \cdots & 0 \\ & & & & & & \ddots & \vdots \\ & & & & & & & 1 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\1\\1\\1\\1\end{matrix}}\right\} m \text{ rows} \\ \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} n-m \text{ rows} \end{matrix} \qquad (m = 1, 2, \cdots, n),$$

where the g's lie in $\mathfrak{T}$, and U_1 is the unit matrix. If

$$\Omega_m = \Omega' U_1 U_2 \cdots U_m = (a_{hk}^{(m)})_{h,k=1,2,\cdots,m} \qquad (m = 1, 2, \cdots, n),$$

then $\Omega_1 = \Omega'$, and for $h, k = 1, 2, \cdots, n$

$$a_{hk}^{(m)} = a_{hk}^{(m-1)} \text{ if } k \neq m, \text{ and } a_{hm}^{(m)} = g_1^{(m)} a_{h1}^{(m-1)} + \cdots + g_{m-1}^{(m)} a_{hm-1}^{(m-1)} + a_{hm}^{(m-1)}.$$

The n principal determinants of Ω_m :

$$\Delta_r = | a_{hk}^{(m)} |_{h,k=1,2,\cdots,r} \qquad (r = 1, 2, \cdots, n)$$

are therefore equal to the corresponding ones of Ω_{m-1} and so of Ω'.

By construction, the elements of Ω_1 satisfy the inequalities

$$| a_{hk}^{(1)} | \leqq \sigma^{(k)} \qquad (h, k = 1, 2, \cdots, n),$$

and therefore also the inequalities

$$| a_{h1}^{(1)} | \leqq \sigma^{(h)} \qquad (h = 1, 2, \cdots, n).$$

Assume now that $U_1, \cdots, U_{m-1}$ were determined such that

$$(39) \qquad \begin{aligned} &| a_{hk}^{(m-1)} | \leqq \sigma^{(k)} && (h, k = 1, 2, \cdots, n); \\ &| a_{hk}^{(m-1)} | \leqq \sigma^{(h)} && \text{for } h = 1, 2, \cdots, n; k = 1, 2, \cdots, m-1. \end{aligned}$$

Then U_m, as we shall prove now, can be constructed such that Ω_m satisfies the stronger inequalities

$$(40) \qquad \begin{aligned} &| a_{hk}^{(m)} | \leqq \sigma^{(k)} && (h, k = 1, 2, \cdots, n); \\ &| a_{hk}^{(m)} | \leqq \sigma^{(h)} && \text{for } h = 1, 2, \cdots, n; k = 1, 2, \cdots, m. \end{aligned}$$

To this purpose put

$$a_{h1}^{(m-1)} \gamma_1 + \cdots + a_{hm-1}^{(m-1)} \gamma_{m-1} + a_{hm}^{(m-1)} = t_h(\gamma_1, \cdots, \gamma_{m-1}) = t_h \quad (h = 1, 2, \cdots, n),$$

and determine elements $\gamma_1, \gamma_2, \cdots, \gamma_{m-1}$ of $\mathfrak{K}$ such that

$$t_1 = t_2 = \cdots = t_{m-1} = 0.$$

This system of linear equations has the determinant Δ_{m-1}. On solving,

$$\Delta_{m-1}\gamma_r = \mp \Delta_{m-1,r} \qquad (r = 1, 2, \cdots, m-1),$$

where $\Delta_{m-1,r}$ is the $(m-1)^{\text{th}}$ order minor of Δ_m obtained by omitting the m^{th} row and the r^{th} column. Hence from (37),

$$| \gamma_r | = \left| \frac{\Delta_{m-1,r}}{\Delta_{m-1}} \right| \leqq \frac{\sigma^{(1)} \cdots \sigma^{(m)}}{\sigma^{(r)}} : (\sigma^{(1)} \cdots \sigma^{(m-1)}) = \frac{\sigma^{(m)}}{\sigma^{(r)}} \geqq 1.$$

Let the element $g_r^{(m)}$ of U_m now be the number in $\mathfrak{T}$ satisfying the inequality

$$| g_r^{(m)} - \gamma_r | < 1 \qquad (r = 1, 2, \cdots, m-1),$$

so that

$$| g_r^{(m)} | = \frac{\sigma^{(m)}}{\sigma^{(r)}}.$$

Then from the first system of inequalities (39) for $h = 1, 2, \cdots, n$

$$| a_{hm}^{(m)} | = | t_h(g_1^{(m)}, \cdots, g_{m-1}^{(m)}) | = | g_1^{(m)} a_{h1}^{(m-1)} + \cdots + g_{m-1}^{(m)} a_{hm-1}^{(m-1)} + a_{hm}^{(m-1)} |$$

$$\leqq \max\left(\frac{\sigma^{(m)}}{\sigma^{(1)}} \cdot \sigma^{(1)}, \cdots, \frac{\sigma^{(m)}}{\sigma^{(m-1)}} \cdot \sigma^{(m-1)}, \sigma^{(m)}\right) = \sigma^{(m)},$$

and from the second system for $h = 1, 2, \cdots, m$

$$| a_{hm}^{(m)} | = | t_h(g_1^{(m)}, \cdots, g_{m-1}^{(m)}) |$$

$$= | (g_1^{(m)} - \gamma_1)a_{h1}^{(m-1)} + \cdots + (g_{m-1}^{(m)} - \gamma_{m-1})a_{hm-1}^{(m-1)} | < 1 \cdot \sigma^{(h)} = \sigma^{(h)}.$$

Since the remaining inequalities (40) are contained in (39), the matrix U_m has the required property. Hence if

$$U = XU_1U_2 \cdots U_n,$$

then this matrix satisfies the statement of our theorem.

13. A property of the product of n inhomogeneous linear polynomials in n variables. Let $\Omega = (a_{hk})_{h,k=1,2,\cdots,n}$ be again a matrix with elements in $\mathfrak{K}$ of determinant 1. We form the distance function

$$F(X \mid f) = \max_{h=1,2,\cdots,n} (e^{f_h} | a_{h1}x_1 + a_{h2}x_2 + \cdots + a_{hn}x_n |),$$

where $f_1, f_2, \cdots, f_n$ are n integers such that $f_1 + \cdots + f_n = 0$. By the theorem of last paragraph, there is a matrix U with elements in $\mathfrak{T}$ and determinant 1, such that the product matrix

$$\Omega^* = \Omega U = (a_{hk}^*)$$

satisfies the equation

$$\prod_{h=1}^{n} \max_{k=1,2,\cdots,n} (| a_{hk}^* |) = 1.$$

Let us choose the integers f_h^0 such that

$$e^{-f_h^0} = \max_{k=1,2,\cdots,n} (| a_{hk}^* |) \qquad (h = 1, 2, \cdots, n) \tag{41}$$

and put

$$a_{hk}^{**} = z^{f_h^0} a_{hk}^* \qquad (h, k = 1, 2, \cdots, n).$$

Then by the transformation $X = UY$, $F(X) \mid f^0)$ changes into a new distance function

$$F(X \mid f^0) = F'(Y) = \max_{h=1,2,\cdots,n} (| a_{h1}^{**} y_1 + \cdots + a_{hn}^{**} y_n |),$$

where now all coefficients a_{hk}^{**} satisfy the inequalities $|a_{hk}| \leqq 1$, and their determinant is still 1. Obviously, for all n $\mathfrak{K}$-independent vectors $Y^{(1)} = E^{(1)}$, $Y^{(2)} = E^{(2)}, \cdots, Y^{(n)} = E^{(n)}$, the value of this function

$$F'(Y^{(h)}) \leqq 1 \qquad (h = 1, 2, \cdots, n).$$

Therefore by the equation (25), necessarily

$$F'(Y^{(1)}) = F'(Y^{(2)}) = \cdots = F'(Y^{(n)}) = 1,$$

and so all minima of $F(X \mid f^0)$, where the f^0's are given by (41), have the same value 1, and in particular, the first minimum of $F(X \mid f^0)$ has the exact value $\frac{1}{\sqrt[n]{V}}$, where $V = 1$ is the volume of $F(X \mid f^0) \leqq 1$.

As an application, let $a_1, a_2, \cdots, a_n$ be any n elements of $\mathfrak{K}$, and $\eta_1, \eta_2, \cdots, \eta_n$ n elements of $\mathfrak{K}$ satisfying the equations

$$a_{h1}^{*}\eta_1 + \cdots + a_{hn}^{*}\eta_h + a_h = 0 \qquad (h = 1, 2, \cdots, n).$$

If $y_1, y_2, \cdots, y_n$ are the elements of $\mathfrak{T}$ for which

$$|y_h - \eta_h| \leqq \frac{1}{e} \qquad (h = 1, 2, \cdots, n),$$

then obviously

$$|a_{h1}^{*}y_1 + \cdots + a_{hn}^{*}y_n + a_h| \leqq e^{-f_h^0 - 1} \quad (h = 1, 2, \cdots, n).$$

Hence the lattice point $X = (x_1, x_2, \cdots, x_n) = U'Y$ *satisfies the inequalities*

$$|a_{h1}x_1 + \cdots + a_{hn}x_n + a_h| \leqq e^{-f_h^0 - 1} \quad (h = 1, 2, \cdots, n),$$

and therefore the inequality

$$\prod_{h=1}^{n} |a_{h1}x_1 + \cdots + a_{hn}x_n + a_h| \leqq e^{-n}.$$

Here the constant e^{-n} on the right-hand side is the best possible, as is clear if, e.g. Ω is the unit matrix and all $a_h = 1/z$.

14. Distance functions in $\mathfrak{K}_\mathfrak{p}$. The field $\mathfrak{R}$ of all rational functions with coefficients in $\mathfrak{k}$ has valuations different from the "infinite" valuation $|x|$, which expresses the behavior of x at the point $z = \infty$.

Let ζ be any element of $\mathfrak{k}$, and $\mathfrak{p}$ the "finite" point $z = \zeta$. Then we define a valuation $|x|_\mathfrak{p}$ by putting for $x \neq 0$

$$|x|_\mathfrak{p} = e^{-f_\mathfrak{p}},$$

where $f_\mathfrak{p}$ is that integer, for which neither the numerator nor the denominator of the simplified fraction $(z - \zeta)^{-f_\mathfrak{p}}x$ are divisible by $z - \zeta$; we denote by $\mathfrak{K}_\mathfrak{p}$

the perfect extension of $\mathfrak{R}$ with respect to this valuation; it consists of all formal Laurent series

$$x = \alpha_f(z - \zeta)^f + \alpha_{f+1}(z - \zeta)^{f+1} + \alpha_{f+2}(z - \zeta)^{f+2} + \cdots$$

with coefficients in $\mathfrak{k}$, and if $\alpha_f \neq 0$, then $| x |_\mathfrak{p} = e^{-f}$.

Let now $F(X)$ be any special distance function of $\mathfrak{K}$; we use it as the measure for the size of X. Further let $F(X \mid \mathfrak{p})$ be a general distance function of $\mathfrak{K}_f$. Since

$$F((z - \zeta)^f X \mid \mathfrak{p}) = e^{-f} F(X \mid \mathfrak{p}),$$

this distance function may assume arbitrarily small values, if X lies in the modul Λ_n of all lattice points. By (5), there is a constant $\Gamma_\mathfrak{p} > 0$ such that

$$F(X \mid \mathfrak{p}) \leqq \Gamma_\mathfrak{p} \, | X |_\mathfrak{p} \, ;$$

here for $X = (x_1, \cdots, x_n)$

$$| X |_\mathfrak{p} = \max \, (| x_1 |_\mathfrak{p}, \cdots, | x_n |_\mathfrak{p}).$$

Hence

$$F(X \mid \mathfrak{p}) \leqq \Gamma_\mathfrak{p} \text{ for all lattice points } X.$$

Let t be an integer such that

$$e^{-t} \leqq \Gamma_\mathfrak{p}, \quad \text{i.e.} \quad t \geqq \log\left(\frac{1}{\Gamma_\mathfrak{p}}\right),$$

and $C(e^{-t} \mid \mathfrak{p})$ the convex set of all points X in P_n for which

$$F(X \mid \mathfrak{p}) \leqq e^{-t}.$$

Then the set $\mathfrak{m}(-t \mid \mathfrak{p})$ of all lattice points in $C(e^{-t} \mid \mathfrak{p})$ contains with X and Y also $aX + bY$, when a and b lie in $\mathfrak{T}$; it is therefore an $\mathfrak{T}$-modul. By the general theory of polynomial ideals,[16] this modul has a basis of n lattice points

$$P^{(k)} = (p_1^{(k)}, \cdots, p_n^{(k)}) \qquad (k = 1, 2, \cdots, n),$$

such that every point X in Λ_n belongs to $\mathfrak{m}(-t \mid \mathfrak{p})$, if and only it can be written as

$$X = y_1 P^{(1)} + \cdots + y_n P^{(n)} \quad \text{with } y_1, \cdots, y_n \text{ in } \mathfrak{T}.$$

The determinant

$$D(-t) = | p_h^{(k)} |_{h,k=1,2,\cdots,n} \neq 0,$$

and therefore the number

$$\Delta(-t) = | D(-t) |$$

is positive.

---

[16] Compare the basis theorem in §80 of van der Waerden's "Moderne Algebra", Vol. II, 1st ed.

The function $F(X)$ changes into a new distance function

$$F'(Y) = F(X) = F(\Omega Y), \qquad \Omega = (p_h^{(k)})_{h,k=1,2,\cdots,n})$$

by the transformation (42). The convex body $F'(Y) \leqq 1$ has the volume

$$V' = \Delta(-t)^{-1}V,$$

where V denotes the volume of $F(X) \leqq 1$. By the results in §9, there are n lattice points $Y^{(1)}, \cdots, Y^{(n)}$ with determinant 1, such that

$$F'(Y^{(1)}) \cdots F(Y^{(n)}) = \frac{\Delta(-t)}{V}.$$

The transformed lattice points $X^{(1)}, \cdots, X^{(n)}$ given by

$$X^{(k)} = \Omega Y^{(k)} = (x_1^{(k)}, \cdots, x_n^{(k)}) \qquad (k = 1, 2, \cdots, n)$$

have the determinant

$$D(-t) = | x_h^{(k)} |_{h,k=1,2,\cdots,n},$$

and satisfy the relations

$$F(X^{(1)}) \cdots F(X^{(n)}) = \frac{\Delta(-t)}{V}, \qquad F(X^{(k)} \mid \mathfrak{p}) \leqq e^{-t} \qquad (k = 1, 2, \cdots, n).$$

It is not difficult to prove that for large t

$$\Delta(-t) = O(e^{nt}), \qquad | D(-t) |_{\mathfrak{p}} = O(e^{-t}).$$

In the following case, sharper results are obtained. Let

$$F(X \mid \mathfrak{p}) = \max_{h=1,2,\cdots,m} (| a_{h1}x_1 + \cdots + a_{hn-m}x_{n-m} + x_{n-m+h} |_{\mathfrak{p}}),$$

where the a's are elements in $\mathfrak{K}_{\mathfrak{p}}$ such that

$$| a_{hk} |_{\mathfrak{p}} \leqq 1 \qquad \begin{pmatrix} h = 1, 2, \cdots, m \\ k = 1, 2, \cdots, n \end{pmatrix}.$$

Then to every positive integer t there are elements A_{hk} in $\mathfrak{T}$ satisfying

$$| a_{hk} - A_{hk} |_{\mathfrak{p}} \leqq e^{-t} \qquad \begin{pmatrix} h = 1, 2, \cdots, m \\ k = 1, 2, \cdots, n \end{pmatrix}.$$

Hence, if $y_1, \cdots, y_n$ belong to $\mathfrak{T}$, and $x_1, \cdots, x_n$ are defined by

$$x_1 = y_1, \cdots, x_{n-m} = y_{n-m};$$

(42) $$x_{n-m+h} = (z - \zeta)^t y_{n-m+h} - (A_{h1}y_1 + \cdots + A_{hn-m}y_{n-m}),$$

$$(h = 1, 2, \cdots, m),$$

then $F(X \mid \mathfrak{p}) \leqq e^{-t}$. Let $F'(Y) = F(X)$ be the special distance function in $\mathfrak{K}$ derived from $F(X)$ by the transformation (42). Then $F'(Y) \leqq 1$ has the

volume $|(z - \zeta)^{-mt}| V = e^{-mt}V$. Hence there are n $\mathfrak{K}$-independent lattice points $Y^{(1)}, \cdots, Y^{(n)}$ of determinant 1 such that

$$F'(Y^{(1)}) \cdots F'Y^{(n)}) = \frac{e^{mt}}{V}.$$

The n lattice points $X^{(1)}, \cdots, X^{(n)}$ derived from these by (42) have the determinant $(z - \zeta)^{mt}$ and satisfy the conditions

$$F(X^{(1)}) \cdots F(X^{(n)}) = \frac{e^{mt}}{V}, \qquad F(X^{(k)} \mid \mathfrak{p}) \leqq e^{-t} \qquad (k = 1, 2, \cdots, n).$$

MANCHESTER, ENGLAND.

Annals of Mathematics
Vol. 42, No. 2, April, 1941

CONCRETE REPRESENTATION OF ABSTRACT (L)-SPACES AND THE MEAN ERGODIC THEOREM

By Shizuo Kakutani

(Received February 6, 1940)

1. **Introduction.**[1] Let Ω be an abstract space where a completely additive measure is defined. As is well known, the totality of all the real-valued measurable functions $x(t)$ which are absolutely integrable on Ω constitutes a Banach space $L(\Omega)$ with $\| x \| = \int_\Omega | x(t) | dt$ as its norm. Although the space $L(\Omega)$ is not necessarily separable,[2] it may always be considered as a semi-ordered Banach space.[3] Indeed, if we denote, for any pair of elements $x(t)$ and $y(t) \in L(\Omega)$, by $x \geqq y$ (or $y \leqq x$) the relation that $x(t) \geqq y(t)$ almost everywhere on Ω, then the following conditions are satisfied ($x, y, z, w \in L(\Omega)$, λ = scalar):

(I) *$x \geqq y$ and $y \geqq x$ imply $x = y$,*[4]

(II) *$x \geqq y$ and $y \geqq z$ imply $x \geqq z$,*

(III) *$x \geqq y$ and $\lambda \geqq 0$ imply $\lambda x \geqq \lambda y$,*

(IV) *$x \geqq y$ implies $x + z \geqq y + z$ for any z,*

(V) *$x_n \geqq y_n$, $x_n \to x$ (strongly) and $y_n \to y$ (strongly) imply $x \geqq y$,*

(VI) *to any pair of elements x and y, there exists a maximum $z = x \vee y$ such that $z \geqq x$, $z \geqq y$, and $z \leqq z'$ for any z' with $z' \geqq x$, $z' \geqq y$,*

(VII) *to any pair of elements x and y, there exists a minimum $w = x \wedge y$ such that $w \leqq x$, $w \leqq y$, and $w \geqq w'$ for any w' with $w' \leqq x$, $w' \leqq y$.*

Moreover, this semi-ordered Banach space $L(\Omega)$ has the following important property:

(VIII) *$x \geqq 0$ and $y \geqq 0$ imply $\| x + y \| = \| x \| + \| y \|$;*

in orther words, calling x to be *positive* in case $x \geqq 0$, norm is additive on positive elements. Such a Banach space was introduced axiomatically by Garrett

[1] The principal results of this paper were previously announced in S. Kakutani [7]. In [7] we have tacitly assumed the condition (IX).

[2] There are two typical cases when $L(\Omega)$ is not separable. The first one is the case of the Haar's measure of a non-separable bicompact topological group, and the second one is the case of the linear measure in the plane. In the first case, the total space Ω is of finite measure and every measurable subset Ω' of Ω with $m(\Omega') > 0$ determines a non-separable Banach space $L(\Omega')$. In the second case, the total space is not expressible as a sum of a countable infinite number of subsets of finite measure, while $L(\Omega')$ is separable for every measurable subset Ω' of Ω with $m(\Omega') < \infty$.

[3] It is to be noted that in the first case (see footnote (2)) there exists an element $x_0 > 0$ (for example, a function $x_0(t)$ which is identically equal to 1) such that $x_0 \wedge x > 0$ for any $x > 0$, while there exists no such element in the second case.

[4] $x = y$ means that we have $x(t) = y(t)$ almost everywhere on Ω.

Birkhoff [3]. He has introduced the space of this type as a generalization of the concrete Banach space (L) (i.e., the space of all the real-valued measurable functions $x(t)$ which are absolutely integrable on $0 \leqq t \leqq 1$), and has discussed the iteration of bounded linear operations in such Banach spaces. We shall call a Banach space with the semi-ordering satisfying (I)–(VIII) an *abstract* (L)-*space* (notation: (AL)). Every Banach space $L(\Omega)$ is an (AL), and, in contrast to general abstract (L)-spaces, this will be called a *concrete* (L)-space.

In the present paper we shall discuss the converse problem, i.e., we shall investigate how it is possible to represent any abstract (L)-space (AL) by a concrete (L)-space $L(\Omega)$. In other words, given an abstract (L)-space (AL) with the semi-ordering satisfying (I)–(VIII), it is required to construct a space Ω and a completely additive measure defined on some Borel field of Ω such that the corresponding Banach space $L(\Omega)$ is equivalent (= isometric and lattice-isomorphic) to the given space (AL).

This problem is not always possible, if we have no further assumptions on (AL). In order to see this, we have only to notice that the property:

(IX) $x \wedge y = 0$ *implies* $\| x + y \| = \| x - y \|$,

which is always satisfied for any concrete (L)-space, does not necessarily follow from the conditions (I)–(VIII). Indeed, if we consider the (x, y)-plane with the usual semi-ordering: $(x_1, y_1) \geqq (x_2, y_2)$ if and only if $x_1 \geqq x_2$ and $y_1 \geqq y_2$ simultaneously, and define its norm by

$$\begin{aligned} \| (x, y) \| &= | x + y | && \text{if} \quad x \geqq 0, \ y \geqq 0 \quad \text{or} \quad x \leqq 0, \ y \leqq 0, \\ &= \sqrt{x^2 + y^2} && \text{if} \quad x \geqq 0, \ y \leqq 0 \quad \text{or} \quad x \leqq 0, \ y \geqq 0, \end{aligned}$$

then the conditions (I)–(VII) are all satisfied, and yet we have $\| (1, 0) + (0, 1) \| = \| (1, 1) \| = 2 > \| (1, 0) - (0, 1) \| = \| (1, -1) \| = \sqrt{2}$.

If, however, the conditions (I)–(IX) are all satisfied, then our problem has a solution. This will be proved in Theorem 7. The proof is divided into three parts (§§3, 4 and 5), and our principal idea is essentially contained in the papers of H. Freudenthal [4] and F. Wecken [14]. Moreover, it is to be noticed that every abstract (L)-space with the properties (I)–(VIII) can be provided with an equivalent norm which satisfies the additional condition (IX) (Theorem 1, §2).

In Theorem 9 (§6), we shall prove a mean ergodic theorem in abstract (L)-spaces. This is a generalization of a result of Garrett Birkhoff [3] and may be considered as one of the most general formulations of the mean ergodic theorem and Markoff's process. It is further to be noted that, by virtue of Theorem 1, the condition (IX) is unnecessary for the validity of this theorem.

In concluding the introduction, we shall list some elementary lemmas concerning the semi-ordered Banach space, which follow directly from the conditions (I)–(VII) and which are needed in the following discussions.

LEMMA 1.1. $\lambda \geqq 0$ *implies* $\lambda(x \vee y) = \lambda x \vee \lambda y$, $\lambda(x \wedge y) = \lambda x \wedge \lambda y$.

LEMMA 1.2. $(x \vee y) + z = (x + z) \vee (y + z)$, $(x \wedge y) + z = (x + z) \wedge (y + z)$.

LEMMA 1.3. $(x \vee y) + (x \wedge y) = x + y$.

LEMMA 1.4. *$x_i \wedge x_j = 0$ ($i \neq j$, $i, j = 1, 2, \cdots, n$) imply $x_1 + x_2 + \cdots + x_n = x_1 \vee x_2 \vee \cdots \vee x_n$.*

LEMMA 1.5. *$x = x_+ - x_-$, where $x_+ = x \vee 0$, $x_- = (-x) \vee 0$ and $x_+ \wedge x_- = 0$.*

LEMMA 1.6. *$(x \vee y) \wedge z = (x \wedge z) \vee (y \wedge z)$, $(x \wedge y) \vee z = (x \vee z) \wedge (y \vee z)$.*

These lemmas will be found in H. Freudenthal [4] and L. Kantorovitch [8].

2. Change of norm.

THEOREM 1. *Every abstract (L)-space (AL) with the semi-ordering satisfying the conditions* (I)–(VIII) *can be provided with an equivalent norm which satisfies the conditions* (I)–(IX).

PROOF: Put $\| x \|^* = \| x_+ \| + \| x_- \|$ for any $x \epsilon (AL)$. Then we have $\| x \| \leqq \| x \|^*$, $\| \lambda x \|^* = |\lambda| \cdot \| x \|^*$ and $\| x + y \|^* \leqq \| x \|^* + \| y \|^*$ for any $x, y \epsilon (AL)$ and $\lambda \gtreqless 0$. The first two relations are almost trivial and the third one may be proved as follows: $\| x + y \|^* = \| (x + y)_+ \| + \| (x + y)_- \| \leqq \| x_+ + y_+ \| + \| x_- + y_- \|$ (since $0 \leqq (x + y)_+ \leqq x_+ + y_+$, $0 \leqq (x + y)_- \leqq x_- + y_-$) $= \| x_+ \| + \| y_+ \| + \| x_- \| + \| y_- \| = \| x \|^* + \| y \|^*$. Thus $\| x \|^*$ may be considered as a norm on (AL). Moreover, as is easily seen, the conditions (V), (VIII) and (IX) are all satisfied for this new norm $\| x \|^*$. Hence all what we have to prove is that the two norms $\| x \|$ and $\| x \|^*$ are equivalent.

In order to show this, denote by $(AL)^*$ the space (AL) metrized by the new norm $\| x \|^*$. We have only to prove that $(AL)^*$ is complete. For, since we have $\| x \| \leqq \| x \|^*$ for any x, the identical transformation: $x \to x$ is a bounded linear transformation which maps $(AL)^*$ biuniquely on (AL). Consequently if $(AL)^*$ is complete, then by a theorem of S. Banach [1] (pp. 40–41), this mapping must be bicontinuous and there exists a constant C such that $\| x \|^* \leqq C \| x \|$ for any x.

Now, in order to prove the completeness of $(AL)^*$, let $\{x_n\}$ ($n = 1, 2, \cdots$) be a fundamental sequence in $(AL)^*$: $\lim_{m,n \to \infty} \| x_m - x_n \|^* = 0$. We have to show that there exists an $\bar{x} \epsilon (AL)^*$ such that $\lim_{n \to \infty} \| x_n - \bar{x} \|^* = 0$. Without the loss of generality we may assume that we have $\| x_m - x_n \|^* \leqq 2^{-n}$ for $m \geqq n$. Since $\| x_m - x_n \| \leqq \| x_m - x_n \|^*$ for any m and n, $\{x_n\}$ ($n = 1, 2, \cdots$) is also a fundamental sequence in (AL), and, by the completeness of (AL), there exists an $\bar{x} \epsilon (AL)$ such that $\lim_{n \to \infty} \| x_n - \bar{x} \| = 0$. We shall show that we have also $\lim_{n \to \infty} \| x_n - \bar{x} \|^* = 0$. For this purpose, put $\bar{x}_{n,p} = x_n \vee x_{n+1} \vee \cdots \vee x_{n+p}$ for $p = 0, 1, 2, \cdots$; $n = 1, 2, \cdots$ ($\bar{x}_{n,0} = x_n$). Then we have $\bar{x}_{n,p} \leqq \bar{x}_{n,p+1} \leqq \bar{x}_{n,p} + (x_{n+p+1} - x_{n+p})_+$ and $\| \bar{x}_{n,p+1} - \bar{x}_{n,p} \| \leqq \| (x_{n+p+1} - x_{n+p})_+ \| \leqq \| x_{n+p+1} - x_{n+p} \|^* \leqq 2^{-(n+p)}$ for $p = 0, 1, 2, \cdots$; $n = 1, 2, \cdots$. Consequently, since $\sum_{p=0}^{\infty} \| \bar{x}_{n,p+1} - \bar{x}_{n,p} \| \leqq \sum_{p=0}^{\infty} 2^{-(n+p)} = 2^{-(n-1)}$, $\lim_{p \to \infty} \bar{x}_{n,p} \equiv \bar{x}_n$ (strongly) exists and this limit $\bar{x}_n$ clearly satisfies $x_m \leqq \bar{x}_n$ and $\| \bar{x}_n - x_m \| \leqq 2^{-(n-1)}$ for $m \geqq n$. Since $\lim \| x_m - \bar{x} \| = 0$ we have $\bar{x} \leqq \bar{x}_n$ and $\| \bar{x}_n - \bar{x} \| \leqq 2^{-(n-1)}$

for $n = 1, 2, \cdots$. Consequently, $\| x_n - \bar{x} \|^* \leqq \| \bar{x}_n - x_n \|^* + \| \bar{x}_n - \bar{x} \|^* = \| \bar{x}_n - x_n \| + \| \bar{x}_n - \bar{x} \| \leqq 2 \cdot 2^{-(n-1)}$ for $n = 1, 2, \cdots$, and thus we have proved that $\lim_{n \to \infty} \| x_n - \bar{x} \|^* = 0$.

The proof of Theorem 1 is now completed.

Thus we have proved that we can introduce in every abstract (L)-space an equivalent norm which satisfies the conditions (I)–(IX). Hence we shall assume hereafter (in §§3, 4 and 5) that all the conditions (I)–(IX) are satisfied.

3. **Direct decomposition.** The principal result of this chapter is stated in Theorem 2. In the case of the space of functions of bounded variation,[5] Theorem 2 has previously been shown by F. Wecken [14].

We begin with elementary lemmas.

LEMMA 3.1. $\| x \vee y - x' \vee y \| \leqq \| x - x' \|$, $\| x \wedge y - x' \vee y \| \leqq \| x - x' \|$. *Consequently, $x_n \to x$ (strongly) implies $x_n \vee y \to x \wedge y$ (strongly) and $x_n \wedge y \to x \wedge y$ (strongly) for any y.*

PROOF. We shall prove only the first relation. $x \vee y = (x' + (x - x')) \vee y \leqq (x' + (x - x')_+) \vee (y + (x - x')_+) = x' \vee y + (x - x')_+$ implies $x \vee y - x' \vee y \leqq (x - x')_+$ and consequently $(x \vee y - x' \vee y)_+ \leqq (x - x')_+$. Analogously, we have $(x \vee y - x' \vee y)_- \leqq (x - x')_-$. Consequently $\| x \vee y - x' \vee y \| = \| (x \vee y - x' \vee y)_+ \| + \| (x \vee y - x' \vee y)_- \| \leqq \| (x - x')_+ \| + \| (x - x')_+ \| = \| x - x' \|$ (by (IX)).

LEMMA 3.2. *$0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n \leqq x_{n+1} \leqq \cdots \leqq y$ implies the existence of $\lim_{n \to \infty} x_n \equiv x'$ (strongly) with $0 \leqq x' \leqq y$.*

PROOF: For each n we have (by (VIII)) $\sum_{i=1}^{n-1} \| x_{i+1} - x_i \| = \| \sum_{i=1}^{n-1} (x_{i+1} - x_i) \| = \| x_n - x_1 \| \leqq \| y - x_1 \|$. Hence $\sum_{i=1}^{\infty} \| x_{i+1} - x_i \| < \infty$ and consequently $\lim_{n \to \infty} x_n = x'$ (strongly) exists and $0 \leqq x' \leqq y$ (by (V)).

LEMMA 3.3. *For any $x \geqq 0$ and $y \geqq 0$, $\lim_{n \to \infty} (nx \wedge y) = P_x(y)$ (strongly) exists and $0 \leqq P_x(y) \leqq y$.*

PROOF: Clear from Lemma 3.3.

LEMMA 3.4. *For any $x \geqq 0$, $y \to P_x(y)$ is a projection operator defined for all $y \geqq 0$.*[6]

$$P_x(y + z) = P_x(y) + P_x(z), \tag{3.1}$$

$$y \leqq z \textit{ implies } P_x(y) \leqq P_x(z), \tag{3.2}$$

$$P_x(\lambda y) = \lambda P_x(y) \textit{ for any } \lambda \geqq 0, \tag{3.3}$$

(3.4) $\| P_x(y) \| \leqq \| y \|$. *More generally, $\| P_x(y) - P_x(y') \| \leqq \| y - y' \|$. Consequently, $y_n \to y$ (strongly) implies $P_x(y_n) \to P_x(y)$ (strongly),*

$$P_x(P_x(y)) = P_x(y). \tag{3.5}$$

[5] This is, indeed, one of the most familiar examples of abstract (L)-spaces. In this case every principal ideal is separable. See also S. Banach and S. Mazur [2].

[6] In order to define $P_x(y)$ for any $y \in (AL)$, we have only to put $P_x(y) = P_x(y_+) - P_x(y_-)$ for any y. It is clear that we have (3.1), (3.3) (for any $\lambda \gtreqless 0$), (3.4) and (3.5) for any y.

PROOF: By Lemma 1.2, $(nx \wedge y) + (nx \wedge z) = (nx + (nx \wedge z)) \wedge (y + (nx \wedge z)) = 2nx \wedge (nx + z) \wedge (y + nx) \wedge (y + z)$. Hence $nx \wedge (y + z) \leqq (nx \wedge y) + (nx \wedge z) \leqq 2nx \wedge (y + z)$. Making $n \to \infty$ we have (3.1). (3.2) is clear since $y \leqq z$ implies $nx \wedge y \leqq nx \wedge z$ for $n = 1, 2, \cdots$. (3.3) is also clear since, by Lemma 1.1, $nx \wedge \lambda y = \lambda\left(\frac{n}{\lambda} x \wedge y\right)$ for $n = 1, 2, \cdots$. The first relation of (3.4) is again clear, and the second one is a direct consequence of the fact that we have $\| nx \wedge y - nx \wedge y' \| \leqq \| y - y' \|$ for $n = 1, 2, \cdots$ (by Lemma 3.1). Lastly, (3.5) follows from the relation that $nx \wedge P_x(y) = \lim_{m \to \infty} (nx \wedge (mx \wedge y))$ (strongly, by Lemma 3.1) $= nx \wedge y$ for $n = 1, 2, \cdots$.

LEMMA 3.5.

(3.6) $$x \leqq x' \textit{ implies } P_x(y) \leqq P_{x'}(y),$$

(3.7) $$x \wedge x' = 0 \textit{ implies } P_x(y) \wedge P_{x'}(y) = 0,$$

(3.8) $$x \wedge x' = 0 \textit{ implies } P_{x+x'}(y) = P_x(y) + P_{x'}(y),$$

(3.9) $x_n \leqq x$ $(n = 1, 2, \cdots)$ *and* $x_n \to x$ *(strongly) imply* $P_{x_n}(y) \to P_x(y)$ *(strongly)*.[7]

PROOF: (3.6) is clear since $x \leqq x'$ implies $nx \wedge y \leqq nx' \wedge y$ for $n = 1, 2, \cdots$. (3.7) is also clear since $x \wedge x' = 0$ implies $(nx \wedge y) \wedge (nx' \wedge y) = 0$ for $n = 1, 2, \cdots$. (3.8) follows from the relation: $n(x + x') \wedge y = (nx + nx') \wedge y = (nx \vee nx') \wedge y$ (by Lemma 1.4) $= (nx \wedge y) \vee (nx' \wedge y)$ (by Lemma 1.6) $= (nx \wedge y) + (nx' \wedge y)$ (by Lemma 1.4, since $(nx \wedge y) \wedge (nx' \wedge y) = 0$) for $n = 1, 2, \cdots$. Lastly we shall prove (3.9): $x_n \leqq x$ implies $mx_n \wedge y \leqq P_{x_n}(y) \leqq P_x(y)$ (by (3.6)) and consequently $\| P_x(y) - P_{x_n}(y) \| \leqq \| P_x(y) - mx_n \wedge y \| \leqq \| P_x(y) - mx \wedge y \| + \| mx \wedge y - mx_n \wedge y \| \leqq \| P_x(y) - mx \wedge y \| + m \| x - x_n \|$ (by Lemma 3.1). Now, for any $\epsilon > 0$ take an m_0 so large that we have $\| P_x(y) - m_0 x \wedge y \| < \epsilon/2$ and then n_0 so large that we have $m_0 \| x - x_n \| < \epsilon/2$ for $n > n_0$. Then we have $\| P_x(y) - P_{x_n}(y) \| < \epsilon$ for $n > n_0$. Since $\epsilon > 0$ is arbitrary, we have $P_{x_n}(y) \to P_x(y)$ (strongly).

LEMMA 3.6. *For any* $x \geqq 0$ *and* $y \geqq 0$, $P_x(y) = 0$ *is equivalent to* $x \wedge y = 0$.

PROOF: Since we have always $P_x(y) \geqq x \wedge y$, $P_x(y) = 0$ implies $x \wedge y = 0$. Conversely, $x \wedge y = 0$ implies $nx \wedge y = 0$ for $n = 1, 2, \cdots$ and consequently $P_x(y) = 0$.

LEMMA 3.7. $x \wedge (y - P_x(y)) = 0$ *for any* $x \geqq 0$ *and* $y \geqq 0$.

PROOF: By (3.5) and (3.1), we have $P_x(y - P_x(y)) + P_x(y) = P_x(y - P_x(y)) + P_x(P_x(y)) = P_x(y)$. Hence $P_x(y - P_x(y)) = 0$ and, by Lemma 3.6, $x \wedge (y - P_x(y)) = 0$.

[7] It is worth noting that $x_n > x$ $(n = 1, 2, \cdots)$ and $x_n \to x$ (strongly) do not necessarily imply $P_{x_n}(y) \to P_x(y)$ (strongly). For example, put $x_n = \frac{1}{n} y$ for $n = 1, 2, \cdots$. Then we have $x_n \geqq 0$ $(n = 1, 2, \cdots)$, $x_n \to x = 0$ (strongly) and yet $P_{x_n}(y) = y$ does not tend to $P_x(y) = P_0(y) = 0$.

DEFINITION 1. For any $x \geqq 0$ and $y \geqq 0$, $x \succ y$ (or $y \prec x$) means that we have $y \wedge u = 0$ for any $u \geqq 0$ with $x \wedge u = 0$.

LEMMA 3.8. *$x \succ P_x(y)$ for any $x \geqq 0$ and $y \geqq 0$.*

PROOF: $x \wedge u = 0$ implies $(nx \wedge y) \wedge u = 0$ for $n = 1, 2, \cdots$, and making $n \to \infty$ we have $P_x(y) \wedge u = 0$.

LEMMA 3.9. *$x \succ y$ is equivalent to $P_x(y) = y$.*

PROOF: By Lemma 3.8, $P_x(y) = y$ implies $x \succ P_x(y) = y$. Conversely, since $x \wedge (y - P_x(y)) = 0$ by Lemma 3.7, $x \succ y$ implies $y - P_x(y) = y \wedge (y - P_x(y)) = 0$ and consequently $y = P_x(y)$.

DEFINITION 2. A set I of positive elements of (AL) is said to be an *ideal* if the following conditions are satisfied:

$$x \in \mathrm{I} \text{ and } y \in \mathrm{I} \text{ imply } x + y \in \mathrm{I}, \tag{3.10}$$

$$x \in \mathrm{I} \text{ and } y \prec x \text{ imply } y \in \mathrm{I}, \tag{3.11}$$

$$x_n \in \mathrm{I}\ (n = 1, 2, \cdots) \text{ and } x_n \to x \text{ (strongly) imply } x \in \mathrm{I}. \tag{3.12}$$

LEMMA 3.10. *For any $x \geqq 0$ the set of all $y \geqq 0$ which satisfy $y \prec x$ is an ideal.*

PROOF: We have only to prove the following three statements:

$$x \succ y \text{ and } x \succ z \text{ imply } x \succ y + z, \tag{3.13}$$

$$x \succ y \text{ and } y \succ z \text{ imply } x \succ z, \tag{3.14}$$

$$x \succ y_n\ (n = 1, 2, \cdots) \text{ and } y_n \to y \text{ (strongly) imply } x \succ y. \tag{3.15}$$

(3.13) is clear since $x \wedge u = 0$ implies $y \wedge u = 0$ and $z \wedge u = 0$, and consequently $0 \leqq (y + z) \wedge u \leqq (y + z) \wedge (u + z) \wedge (y + u) \wedge (u + u) = y \wedge u + z \wedge u$ (by Lemma 1.2) $= 0$. (3.14) is also clear since $x \wedge u = 0$ implies $y \wedge u = 0$ (since $x \succ y$) and this again implies $z \wedge u = 0$ (since $y \succ z$). Lastly, (3.15) follows from Lemma 3.1. For, $y_n \wedge u = 0$ $(n = 1, 2, \cdots)$ and $y_n \to y$ (strongly) imply $y \wedge u = \lim_{n\to\infty} (y_n \wedge u) = 0$ (by Lemma 3.1).

DEFINITION 3. The ideal obtained in Lemma 3.10 is called the *principal ideal* with *unit* x and is denoted by $[x]$. It is clear that $y \in [x]$ and $y > 0$ implies $y \wedge x > 0$. For each principal ideal, the unit is not unique. For example, we have $[x] = [\lambda x]$ for any $\lambda > 0$. More generally, any $y \geqq 0$ which satisfies $x \succ y$ and $y \succ x$ simultaneously has the property: $[x] = [y]$.

The totality of all the positive elements of (AL) constitutes itself an ideal. We shall call this ideal a unit ideal. The unit ideal is not necessarily principal, and we have

THEOREM 2. *The unit ideal is decomposed into a direct sum of a (not necessarily countable) number of principal ideals. More precisely, there exists a family of principal ideals $\{[x_\alpha]\}$ $(\alpha \in \mathfrak{M})$ such that $x_\alpha \wedge x_\beta = 0$ for any $\alpha \neq \beta$, and any $y > 0$ can be uniquely expressed in the form: $y = \sum_{n=1}^{\infty} y_{\alpha_n}$, $y_{\alpha_n} \in [x_{\alpha_n}]$, where $\{\alpha_n\}$ $(n = 1, 2, \cdots)$ is a countable sequence of indices from $\mathfrak{M}$ which depends on y such that $P_{x_{\alpha_n}}(y) = y_{\alpha_n}$ $(n = 1, 2, \cdots)$ and $P_{x_\alpha}(y) = 0$ for other $\alpha \in \mathfrak{M}$.*

PROOF: Let all the positive elements of (AL) be arranged in the well-ordered sequence: $z_0, z_1, z_2, \cdots, z_\alpha, \cdots, \alpha < \varphi$. We shall define the transfinite sequence $\{x_\alpha\}$, $\alpha < \varphi$, by transfinite induction. Put $x_0 = z_0$, and assume that x_β is already defined for all $\beta < \alpha$. Then put $x_\alpha = z_\alpha$ if $z_\alpha \wedge x_\beta = 0$ for any $\beta < \alpha$, and $x_\alpha = 0$ otherwise (i.e., if there exists at least one $\beta < \alpha$ with $z_\alpha \wedge x_\beta > 0$). In this way, x_α can be defined for all $\alpha < \varphi$. It is clear that we have $x_\alpha \wedge x_\beta = 0$ for any $\alpha \neq \beta$, and that for any $x > 0$ there exists at least one $\alpha < \varphi$ with $x_\alpha \wedge x > 0$. We shall show that this sequence $\{x_\alpha\}$ ($\alpha < \varphi$, omitting those α with $x_\alpha = 0$, and this set of indices will be denoted by $\mathfrak{M}$) is the required one.

In order to prove this, consider, for any $y > 0$, the set $\{P_{x_\alpha}(y)\}$ ($\alpha \in \mathfrak{M}$). Since $0 \leqq P_{x_\alpha}(y) \leqq y$ for any $\alpha \in \mathfrak{M}$ and since $P_{x_\alpha}(y) \wedge P_{x_\beta}(y) = 0$ for any $\alpha \neq \beta$ (by (3.7)), we have $\sum_{i=1}^{n} P_{x_{\alpha_i}}(y) = P_{x_{\alpha_1}}(y) \vee P_{x_{\alpha_2}}(y) \vee \cdots \vee P_{x_{\alpha_n}}(y) \leqq y$ for any finite system of indices $\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ from $\mathfrak{M}$ (by Lemma 1.4). Hence the set of indices $\alpha \in \mathfrak{M}$ with $P_{x_\alpha}(y) > 0$ is at most countable,[8] and if we denote these by $\{\alpha_n\}$ ($n = 1, 2, \cdots$), then the strong limit $\lim_{n\to\infty} \sum_{i=1}^{n} P_{x_{\alpha_i}}(y) \equiv \sum_{n=1}^{\infty} P_{x_{\alpha_n}}(y)$ exists and $\sum_{n=1}^{\infty} P_{x_{\alpha_n}}(y) \leqq y$. We shall prove that this is an equality. Indeed, if we have $y' \equiv y - \sum_{n=1}^{\infty} P_{x_{\alpha_n}}(y) > 0$, then there must exist at least one index $\alpha \in \mathfrak{M}$ with $x_\alpha \wedge y' > 0$. This is, however, a contradiction since we have $0 \leqq x_\alpha \wedge y' \leqq x_\alpha \wedge (y - P_{x_\alpha}(y)) = 0$ for any $\alpha \in \mathfrak{M}$ (by Lemma 3.6).

Thus we have proved that there exists a sequence of indices $\{\alpha_n\}$ ($n = 1, 2, \cdots$) from $\mathfrak{M}$ such that $y = \sum_{n=1}^{\infty} P_{x_{\alpha_n}}(y)$ and $P_{x_\alpha}(y) = 0$ for other $\alpha \in \mathfrak{M}$. In order to prove the uniqueness of this expression, let us assume that we have $y = \sum_{n=1}^{\infty} y_{\beta_n}$, $0 < y_{\beta_n} \in [x_{\beta_n}]$ ($n = 1, 2, \cdots$). Then we have $y \geqq y_{\beta_n}$ and consequently $P_{x_{\beta_n}}(y) \geqq P_{x_{\beta_n}}(y_{\beta_n}) = y_{\beta_n} > 0$ (by Lemma 3.9) for $n = 1, 2, \cdots$. Hence $\{\beta_n\}$ ($n = 1, 2, \cdots$) is a subsequence of $\{\alpha_n\}$ ($n = 1, 2, \cdots$). We shall prove that the totality of $\{\beta_n\}$ ($n = 1, 2, \cdots$) coincides with $\{\alpha_n\}$ ($n = 1, 2, \cdots$) and that we have $P_{x_{\beta_n}}(y) = y_{\beta_n}$ for $n = 1, 2, \cdots$. For, otherwise, we should have $y = \sum_{n=1}^{\infty} P_{x_{\alpha_n}}(y) > \sum_{n=1}^{\infty} P_{x_{\beta_n}}(y) \geqq \sum_{n=1}^{\infty} y_{\beta_n} = y$ or $y = \sum_{n=1}^{\infty} P_{x_{\alpha_n}}(y) \geqq \sum_{n=1}^{\infty} P_{x_{\beta_n}}(y) > \sum_{n=1}^{\infty} y_{\beta_n} = y$, which is clearly a contradiction.

This concludes the proof of Theorem 2.

THEOREM 3. *In order that the unit ideal is principal, it is necessary and sufficient that $\mathfrak{M}$ is at most countable.*

PROOF: If there exists a unit $\mathbf{1}$ such that $\mathbf{1} > x$ for any $x > 0$, or equivalently, $\mathbf{1} \wedge x > 0$ for any $x > 0$, then we have $x'_\alpha \equiv \mathbf{1} \wedge x_\alpha > 0$ for any $\alpha \in \mathfrak{M}$. Since $x'_\alpha \wedge x'_\beta = 0$ for any $\alpha \neq \beta$, we have $0 < x'_{\alpha_1} + x'_{\alpha_2} + \cdots + x'_{\alpha_n} = x'_{\alpha_1} \vee x'_{\alpha_2} \vee \cdots \vee x'_{\alpha_n} \leqq \mathbf{1}$ for any finite system of indices $\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ from $\mathfrak{M}$ (by Lemma 1.4). Hence $\mathfrak{M}$ must be at most countable.[9] Conversely, if $\mathfrak{M}$ is

[8] We have only to notice that for each n there exists only a finite number of indices α such that $\| P_{x_\alpha}(y) \| > 1/n$.

[9] Analogously as in footnote 8. We have only to notice that for each n there exists only a finite number of indices α such that $\| x'_\alpha \| > 1/n$.

at most countable: $\mathfrak{M} = \{\alpha_1, \alpha_2, \cdots\}$, then $\mathbf{1} \equiv \sum_{n=1}^{\infty} 2^{-n} \| x_{\alpha_n} \|^{-1} \cdot x_{\alpha_n}$ will be a unit. Indeed, $x > 0$ implies the existence of an index α_n with $x_{\alpha_n} \wedge x > 0$ and consequently $\mathbf{1} \wedge x \geqq 2^{-n} \cdot \| x_{\alpha_n} \|^{-1} \cdot x_{\alpha_n} \wedge x > 0$.

THEOREM 4. *Every separable abstract (L)-space has a unit.*

PROOF: If we put $x'_\alpha = \| x_\alpha \|^{-1} \cdot x_\alpha$ for any $\alpha \in \mathfrak{M}$, then we have $\| x'_\alpha \| = 1$ and $x'_\alpha \wedge x'_\beta = 0$ for any $\alpha \neq \beta$. Hence (by (IX)) we have $\| x'_\alpha - x'_\beta \| = \| x'_\alpha + x'_\beta \| = \| x'_\alpha \| + \| x'_\beta \| = 2$ for any $\alpha \neq \beta$. Since (AL) is separable by assumption, $\mathfrak{M}$ must be at most countable and, by Theorem 3, the unit ideal of (AL) must be principal.

REMARK. This result may also be obtained directly as follows: Let $\{x_n\}$ $(n = 1, 2, \cdots)$ be a countable set which is dense in the positive part of (AL). If we put $\mathbf{1} \equiv \sum_{n=1}^{\infty} 2^{-n} \cdot \| x_n \|^{-1} \cdot x_n$, then $\mathbf{1}$ is a unit of (AL). Indeed, for any $x > 0$ there exists a subsequence $\{x_{n_\nu}\}$ $(\nu = 1, 2, \cdots)$ of $\{x_n\}$ $(n = 1, 2, \cdots)$ such that $x_{n_\nu} \to x$ (strongly) and $x_{n_\nu} \wedge x \to x \wedge x$ (strongly) $= x > 0$. Hence $x_{n_\nu} \wedge x > 0$ for some ν, and this implies $\mathbf{1} \wedge x \geqq 2^{-n_\nu} \cdot \| x_{n_\nu} \|^{-1} \cdot x_{n_\nu} \wedge x > 0$.

This result was also obtained by H. Freudenthal [4].

4. **Integral representation.** In §3 we have obtained a direct decomposition of the positive part of (AL) into principal ideals. Consequently, our problem of concrete representation is reduced to the case of a principal ideal, i.e., the case when the unit element $\mathbf{1}$ exists.[10] In this chapter we shall show that *the positive part of an abstract (L)-space with unit may be represented by an integral in some abstract Boolean algebra with unit.* This may be considered as a generalization of a well-known result of O. Nikodym [9] (see also S. Saks [10]), and is essentially contained in the paper of H. Freudenthal [4]. The proof given below, however, has some interest.

Let us denote by $\mathbf{1}$ the unit element which we assume to exist (throughout this chapter). Without loss of generality, we may assume that $\| \mathbf{1} \| = 1$.

DEFINITION 4.[11] A positive element $e \geqq 0$ is said to be a *characteristic element* of (AL) if we have $e \wedge (\mathbf{1} - e) = 0$, or equivalently by Lemma 1.2, $2e \wedge \mathbf{1} = e$. The totality of all characteristic elements of (AL) will be denoted by $\mathbf{E}$. It is clear that $e \in \mathbf{E}$ implies $0 \leqq e \leqq \mathbf{1}$.

LEMMA 4.1. *$P_x(\mathbf{1}) \in \mathbf{E}$ for any $x \geqq 0$.*

PROOF: $2P_x(\mathbf{1}) \wedge \mathbf{1} = \lim_{n \to \infty} (2(nx \wedge \mathbf{1}) \wedge \mathbf{1})$ (strongly by Lemma 3.1) $= \lim_{n \to \infty} (2nx \wedge \mathbf{1}) = P_x(\mathbf{1})$.

LEMMA 4.2. *$e \in \mathbf{E}$ is equivalent to $P_e(\mathbf{1}) = e$.*

PROOF: By Lemma 4.1, $P_e(\mathbf{1}) = e$ implies $e \in \mathbf{E}$. Conversely, $e \in \mathbf{E}$ is equivalent to $2e \wedge \mathbf{1} = e$ by definition. And if $2^n e \wedge \mathbf{1} = e$, then $2^{n+1} e \wedge \mathbf{1} = 2(2^n e \wedge \mathbf{1}) \wedge \mathbf{1} = 2e \wedge \mathbf{1} = e$. Hence, $P_e(\mathbf{1}) = \lim_{n \to \infty} (2^n e \wedge \mathbf{1}) = e$.

LEMMA 4.3. *$e \in \mathbf{E}$ implies $\mathbf{1} - e \in \mathbf{E}$.* (Clear.)

[10] See the last lines of the proof of Theorem 7.

[11] Cf. H. Freudenthal [4].

LEMMA 4.4. *$e_1 \in \mathbf{E}$ and $e_2 \in \mathbf{E}$ imply $e_1 \vee e_2 \in \mathbf{E}$ and $e_1 \wedge e_2 \in \mathbf{E}$.*

PROOF: $2(e_1 \vee e_2) \wedge \mathbf{1} = (2e_1 \vee 2e_2) \wedge \mathbf{1} = (2e_1 \wedge \mathbf{1}) \vee (2e_2 \wedge \mathbf{1})$ (by Lemma 1.6) $= e_1 \vee e_2$. $2(e_1 \wedge e_2) \wedge \mathbf{1} = (2e_1 \wedge 2e_2) \wedge \mathbf{1} = (2e_1 \wedge \mathbf{1}) \wedge (2e_2 \wedge \mathbf{1}) = e_1 \wedge e_2$.

LEMMA 4.5. *$e_n \in \mathbf{E}$ $(n = 1, 2, \cdots)$ and $e_n \to e$ (strongly) imply $e \in \mathbf{E}$.*

PROOF: $2e \wedge \mathbf{1} = \lim_{n\to\infty} (2e_n \wedge \mathbf{1})$ (by Lemma 2.1) $= \lim_{n\to\infty} e_n = e$.

LEMMA 4.6. *$\mathbf{E}$ is a Boolean algebra with $e_1 \vee e_2$, $e_1 \wedge e_2$ and $\mathbf{1} - e$ as its fundamental operations, and is closed in (AL) in the strong topology.*

PROOF: Clear from Lemmas 1.6, 4.3, 4.4 and 4.5.

THEOREM 5. *For any $x \geqq 0$ there exists a system of characteristic elements $\{e(\lambda)\}$ $(0 \leqq \lambda < \infty)$, called the resolution of unity, such that*

$$\lambda \leqq \mu \text{ implies } e(\lambda) \leqq e(\mu), \tag{4.1}$$

$$\lambda_n \leqq \lambda \ (n = 1, 2, \cdots) \text{ and } \lambda_n \to \lambda \text{ imply } e(\lambda_n) \to e(\lambda) \text{ (strongly)}, \tag{4.2}$$

$$e(0) = 0, \lim_{\lambda\to\infty} e(\lambda) \equiv e(\infty) = \mathbf{1}, \tag{4.3}$$

$$e \in \mathbf{E} \text{ and } e \leqq e(\lambda) \text{ imply } P_e(x) \leqq \lambda e, \tag{4.4}$$

$$e \in \mathbf{E} \text{ and } e \leqq \mathbf{1} - e(\lambda) \text{ imply } P_e(x) \geqq \lambda e. \tag{4.5}$$

PROOF: Put $e(\lambda) = P_{(\lambda\mathbf{1}-x)_+}(\mathbf{1})$. Then $e(\lambda)$ is characteristic by Lemma 4.1. Since $(\lambda\mathbf{1} - x)_+ \leqq (\mu\mathbf{1} - x)_+$ for $\lambda \leqq \mu$, (4.1) is a direct consequence of (3.6). Analogously, (4.2) is a direct consequence of (3.9), if we observe that we have $(\lambda_n\mathbf{1} - x)_+ = (\lambda_n\mathbf{1} - x) \vee 0 \to (\lambda\mathbf{1} - x) \vee 0 = (\lambda\mathbf{1} - x)_+$ (strongly by Lemma 3.1). The first part of (4.3) is almost evident; for, we have $e(0) = P_{(-x)_+}(\mathbf{1}) = P_0(\mathbf{1}) = 0$. Before coming to the proof of the second part, we shall prove (4.4) and (4.5). $e \in \mathbf{E}$ and $e \leqq e(\lambda)$ imply $0 \leqq e \wedge (x - \lambda\mathbf{1})_+ \leqq e(\lambda) \wedge (x - \lambda\mathbf{1})_+ = \lim_{n\to\infty} (n(\lambda\mathbf{1} - x)_+ \wedge \mathbf{1} \wedge (x - \lambda\mathbf{1})_+)$ (strongly by Lemma 3.1) $= 0$ (by Lemma 1.5), and consequently $P_e((x - \lambda\mathbf{1})_+) = 0$. Hence the trivial relation $x \leqq (x - \lambda\mathbf{1})_+ + \lambda\mathbf{1}$ implies $P_e(x) \leqq P_e((x - \lambda\mathbf{1})_+) + P_e(\lambda\mathbf{1}) = P_e(\lambda\mathbf{1}) = \lambda e$ (by (3.2), (3.1), (3.3) and Lemma 4.2). In the same manner, $e \in \mathbf{E}$ and $e \leqq \mathbf{1} - e(\lambda)$ imply $0 \leqq e \wedge (\lambda\mathbf{1} - x)_+ \leqq (\mathbf{1} - e(\lambda)) \wedge (\lambda\mathbf{1} - x)_+ = (\mathbf{1} - P_{(\lambda\mathbf{1}-x)_+}(\mathbf{1})) \wedge (\lambda\mathbf{1} - x)_+ = 0$ (by Lemma 3.7), and consequently $P_e((\lambda\mathbf{1} - x)_+) = \lim_{n\to\infty} (ne \wedge (\lambda\mathbf{1} - x)_+)$ (strongly) $= 0$. Hence the trivial relation $x + (\lambda\mathbf{1} - x)_+ \geqq \lambda\mathbf{1}$ implies $P_e(x) = P_e(x) + P_e((\lambda\mathbf{1} - x)_+) \geqq P_e(\lambda\mathbf{1}) = \lambda e$. Thus (4.4) and (4.5) are proved.

Now, in order to prove the second relation of (4.3), put $\lim_{\lambda\to\infty} e(\lambda) = e(\infty)$ (strongly), which surely exists and belongs to $\mathbf{E}$ by Lemma 4.5. If we further put $e'(\infty) = \mathbf{1} - e(\infty)$, then $e'(\infty) \in \mathbf{E}$ by Lemma 4.3. Since $e'(\infty) \leqq \mathbf{1} - e(\lambda)$ for any $\lambda > 0$, we have $P_{e'(\infty)}(x) \geqq \lambda e'(\infty)$ and consequently $\| P_{e'(\infty)}(x) \| \geqq \lambda \| e'(\infty) \|$ for any $\lambda > 0$. From this follows directly that we have $e'(\infty) = 0$ and $e(\infty) = \mathbf{1}$.

Thus the proof of Theorem 5 is completed.

THEOREM 6. *Each positive element $x \geqq 0$ of (AL) can be expressed in the form:*

$$x = \int_0^\infty \lambda \, de(\lambda), \tag{4.6}$$

where the integration is of abstract Radon-Stieltjes type and $\{e(\lambda)\}$ $(0 \leqq \lambda < \infty)$ is the resolution of unity obtained in Theorem 5.

PROOF: For any division Δ: $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_n = \Lambda < \infty$ of the interval $(0, \Lambda)$ with $0 < \lambda_i - \lambda_{i-1} < \epsilon$ $(i = 1, 2, \cdots, n)$, we have

$$m(\Delta) \equiv \sum_{i=1}^{n} \lambda_{i-1}(e(\lambda_i) - e(\lambda_{i-1})) \leqq \sum_{i=1}^{n} P_{(e(\lambda_i)-e(\lambda_{i-1}))}(x) = P_{e(\Lambda)}(x)$$

$$\leqq \sum_{i=1}^{n} \lambda_i(e(\lambda_i) - e(\lambda_{i-1})) \equiv M(\Delta)$$

and

$$M(\Delta) - m(\Delta) = \sum_{i=1}^{n} (\lambda_i - \lambda_{i-1})(e(\lambda_i) - e(\lambda_{i-1}))$$

$$\leqq \epsilon \sum_{i=1}^{n} (e(\lambda_i) - e(\lambda_{i-1})) = \epsilon(e(\Lambda) - e(0)) \leqq \epsilon\, \mathbf{1}.$$

These relations follow directly from the fact that we have $e(\lambda_i) - e(\lambda_{i-1}) \leqq e(\lambda_i)$ and $e(\lambda_i) - e(\lambda_{i-1}) \leqq \mathbf{1} - e(\lambda_{i-1})$ for $i = 1, 2, \cdots, n$. Hence we have (by making $\epsilon \to 0$) $P_{e(\Lambda)}(x) = \int_0^\Lambda \lambda \, de(\lambda)$, and, by making $\Lambda \to \infty$, we have the required relation (4.6) (since $\Lambda \to \infty$ implies $e(\Lambda) \to \mathbf{1}$ (strongly) and $P_{e(\Lambda)}(x) \to P_{\mathbf{1}}(x) = x$ (strongly) by (3.9)).

The proof of Theorem 6 is hereby completed.

REMARK. Theorems 5 and 6 are also valid even if there exists no unit element in (AL); for, we have only to consider the principal ideal $[x]$.

5. **Concrete representation.** In §4 we have seen that any positive element of an (AL) with unit can be represented by an integral in some abstract Boolean algebra. We shall show, in this chapter, that this abstract Boolean algebra (with a unit element) can be represented by a concrete one with a completely additive measure, and that the abstract integration can be substituted by a concrete one.

THEOREM 7. *To any abstract (L)-space (AL) satisfying the conditions* (I)–(IX), *with a unit element there corresponds a totally disconnected (bicompact) topological space Ω and a completely additive measure defined on a Borel field of Ω, such that (AL) is isometric and lattice-isomorphic to the Banach space $L(\Omega)$.*

PROOF: We shall first treat the case when the unit exists. By Lemma 4.6, the totality $\mathbf{E}$ of all the characteristic elements e of (AL) constitutes a Boolean algebra with $\mathbf{1}$ as its unit element. Hence, by a theorem of M. H. Stone [12]-

H. Wallman [13], **E** may be represented by a concrete Boolean algebra **K** of all the simultaneously open and closed subsets of a totally disconnected bicompact topological space Ω. Let E be the element of **K** which corresponds to the element e of **E**. Then $m(E) = \| e \|$ is clearly a finitely additive measure defined on **K**, and, if we put $d(E_1, E_2) = m(E_1 + E_2 - E_1E_2)$, then (denoting by e_1 and e_2 the corresponding elements in **E**), $d(E_1, E_2) = m(E_1 + E_2 - E_1E_2) = m((E_1 + E_2) - E_2) + m((E_1 + E_2) - E_1) = \| e_1 \vee e_2 - e_2 \| + \| e_1 \vee e_2 - e_1 \| = \| (e_1 - e_2) \vee 0 \| + \| 0 \vee (e_2 - e_1) \| = \| e_1 - e_2 \|$ (by IX)); i.e., with this metric $d(E_1, E_2)$ **E** is isometric with **K**.

We shall next prove that this $m(E)$ is completely additive on **K**. For this purpose, it suffices to show that no element E of **K** can be expressed as a sum of a countable infinite number of non-vacuous disjoint sets $\{E_n\}$ $(n = 1, 2, \cdots)$ of **K**. Indeed, if we have $E = \sum_{n=1}^{\infty} E_n$, $E \in \mathbf{K}$, $E_n \in \mathbf{K}$ $(n = 1, 2, \cdots)$ and $E_mE_n = 0$ $(m \neq n)$, then the closed set E is covered by a system of open sets $\{E_n\}$ $(n = 1, 2, \cdots)$. Since the space Ω is bicompact, E is covered by a finite number of E_n, and this is clearly impossible since each E_n $(n = 1, 2, \cdots)$ is non-vacuous and $E_m \cdot E_n = 0$ for $m \neq n$.

Thus we have proved that $m(E)$ is completely additive on **K**. Hence, by a theorem of E. Hopf [5] (p. 2), $m(E)$ can be extended to the least Borel field $B(\mathbf{K})$ containing **K**, and it will be easily seen that the residual class of $B(\mathbf{K})$ modulo the ideal of all the sets of measure zero of $B(\mathbf{K})$ forms a Boolean algebra which is isometric and lattice-isomorphic to **E**.[12] Thus we have obtained a space Ω and a completely additive measure $m(E)$ defined on the Borel field $B(\mathbf{K})$ of subsets E of Ω which is isometric and lattice-isomorphic to **E** (if we neglect the sets of measure zero); and it is now an easy matter to show (under the condition (IX)) that (AL) is isometric and lattice-isomorphic to the Banach space $L(\Omega)$ which is determined by the measure just obtained above. Indeed, the integral representation obtained in Theorem 6 is simply the one-to-one norm-preserving correspondence of the positive part of (AL) and $L(\Omega)$, and from this follows (by virtue of (IX)) that (AL) and $L(\Omega)$ are isometric and lattice-isomorphic to each other.

Thus Theorem 7 is proved under the condition that the unit element exists. In order to discuss the general case, we have only to appeal to Theorem 2. Indeed, we have only to consider the family of spaces $\{\Omega_\alpha\}$, $\alpha \in \mathfrak{M}$, corresponding to the principal ideals $\{[x_\alpha]\}$, $\alpha \in \mathfrak{M}$, and to put $\Omega = \sum_{\alpha \in \mathfrak{M}} \Omega_\alpha$. As is easily

[12] For this purpose we have only to show that for any $E \in B(\mathbf{K})$ with $E = \sum_{n=1}^{\infty} E_n, E_n \in \mathbf{K}$ $(n = 1, 2, \cdots)$ and $E_1 \subset E_2 \subset \cdots \subset E_n \subset \cdots$, there exists an $E' \in \mathbf{K}$ such that $E' \supset E$ and $m(E' - E) = 0$. In order to show this, consider the corresponding elements $\{e_n\}$ $(n = 1, 2, \cdots)$ from **E**. Then we have $e_1 \leqq e_2 \leqq \cdots \leqq e_n \leqq \cdots$, and $\| e_n \| = m(E_n) \leqq m(E) < \infty$. Hence $\lim_{n \to \infty} e_n = e'$ exists and, if we denote by E' the element of **K** which corresponds to e', then E' satisfies $E' \supset E_n$ $(n = 1, 2, \cdots)$ and $m(E') = \| e' \| = \lim_{n \to \infty} \| e_n \| = \lim_{n \to \infty} m(E_n)$. Consequently $E' \supset \sum_{n=1}^{\infty} E_n = E$ and $m(E' - E) = \lim m(E' - E_n) = 0$.

seen, $L(\Omega)$ is the totality of all the measurable functions $x(t)$ defined all over $\Omega = \sum_{\alpha\in\mathfrak{A}} \Omega_\alpha$, for which there exists at most a countable number of indices $\{\alpha_n\}$ $(n = 1, 2, \cdots)$ (depending on $x(t)$) such that we have $\sum_{n=1}^{\infty} \int_{\Omega_{\alpha_n}} |x(t)| \, dt < \infty$ and $x(t) = 0$ almost everywhere on other Ω_α. It is clear that (AL) and $L(\Omega)$ are isometric and lattice-isomorphic to each other.

The proof of Theorem 7 is now completed.

If, moreover, the space (AL) is separable, then $B(\mathbf{K})$ is also separable with respect to the metric $d(E_1, E_2) = m(E_1 + E_2 - E_1E_2)$. Furthermore, by Theorem 4, this (AL) has a unit $\mathbf{1}$ with $\|\mathbf{1}\| = 1$. Consequently $B(\mathbf{K})$ has a unit element $\mathbf{1}$ with $m(\mathbf{1}) = 1$. Hence, by a well-known argument, $B(\mathbf{K})$ can be embedded isometrically and lattice-isomorphically in a Boolean algebra of all the measurable sets of the unit interval $(0, 1)$.[13]

THEOREM 8. *Every separable abstract (L)-space (satisfying* (I)–(IX)) *can be embedded isometrically and lattice-isomorphically into the Banach space (L) (= the concrete (L)-space defined on the unit interval* $(0, 1)$ *with respect to the ordinary Lebesgue measure).*

6. Mean ergodic theorem in abstract (L)-spaces.

DEFINITION 5. Let T be a bounded linear operation which maps a semi-ordered Banach space into itself. T is said to be *positive* if we have $T(x) \geqq 0$ for any $x \geqq 0$.

THEOREM 9. *Let T be a positive bounded linear operation which maps an abstract (L)-space (AL) (satisfying the conditions* (I)–(VIII)) *into itself. If there exists a constant C such that $\| T^n \| \leqq C$ for $n = 1, 2, \cdots$, and if there exists for any $x \geqq 0$ an $x_0 \geqq 0$ such that $x_n \equiv \frac{1}{n}(x + T(x) + \cdots T^{n-1}(x)) \leqq x_0$ for $n = 1, 2, \cdots$, then for any $x \in (AL)$ the sequence $\{x_n\}$ $(n = 1, 2, \cdots)$ converges strongly to a point $\bar{x} \in (AL)$, i.e., mean ergodic theorem is valid in (AL).*

REMARK. This is a generalization of a result of Garrett Birkhoff [3]. He has assumed that T preserves the norm of positive elements (i.e., $\| T(x) \| = \| x \|$ for any $x \geqq 0$), and has only proved that $\{f(x_n)\}$ $(n = 1, 2, \cdots)$ converges for any bounded linear functional $f(x)$ defined on (AL). (Since the *weak completeness* of (AL) (see Theorem 11) was not proved by him, the weak convergence of $\{x_n\}$ $(n = 1, 2, \cdots)$ was not yet established).

PROOF OF THEOREM 9. By Theorem 1, we have only to discuss the case when the conditions (I)–(IX) are all satisfied. Hence we shall assume, appealing to Theorem 7, that (AL) is represented isometrically and lattice-isomorphically by a concrete (L)-space $L(\Omega)$. Ω may not be a sum of a countable infinite number of subsets of finite measure. By a theorem of K. Yosida [15] and the author [7] (see also F. Riesz [10]), we have only to prove that, for any $x \in L(\Omega)$, the se-

[13] Cf. S. Bochner and J. v. Neumann, *On compact solutions of operational-differential equations I*, Annals of Math., Vol. 36 (1935), p. 264, Footnote 17.

quence $\{x_n\}$ $(n = 1, 2, \cdots)$ contains a subsequence which converges weakly to some $\bar{x} \in L(\Omega)$. Since any $x \in L(\Omega)$ can be represented as a difference of two positive elements: $x = x_+ - x_-$, we have only to discuss the case $x \geqq 0$, and our theorem is reduced to the following

THEOREM 10. *Let $\{x_n(t)\}$ $(n = 1, 2, \cdots)$ be a sequence of non-negative measurable functions from $L(\Omega)$. (Ω may not be a sum of a countable infinite number of subsets of finite measure.) If there exists a function $x_0(t) \in L(\Omega)$ such that $x_n(t) \leqq x_0(t)$ almost everywhere on Ω, then the sequence $\{x_n(t)\}$ $(n = 1, 2, \cdots)$ contains a subsequence which converges weakly to some function $\bar{x}(t) \in L(\Omega)$.*

PROOF: When Ω is the interval $(0, 1)$ or $(-\infty, +\infty)$, this theorem is well-known, and the general case can be reduced to this case. To show this, consider the family $\mathbf{K}$ of all the sets $E_{n,\alpha} = E_t[x_n(t) > \alpha]$, where α is any positive number and $n = 0, 1, 2, \cdots$. Each $E_{n,\alpha}$ is of finite measure since we have $\alpha m(E_{n,\alpha}) \leqq \int_\Omega x_n(t)\, dt < \infty$ for any $\alpha > 0$ and $n = 0, 1, 2, \cdots$. Moreover, since we have $E_{n,\alpha} = \lim_{k\to\infty} E_{n,r_k}$, where α is any positive number and $\{r_k\}$ $(k = 1, 2, \cdots)$ is a monotone decreasing sequence of positive rational numbers which tends to α, $\mathbf{K}$ is separable with respect to the metric $d(E_1, E_2) = m(E_1 + E_2 - E_1E_2)$. Hence, by a well-known argument, $\mathbf{K}$ may be represented isometrically and lattice-isomorphically by a family of measurable sets in the interval $(-\infty, +\infty)$.

Thus the characteristic functions of the sets of $\mathbf{K}$ and consequently the functions $\{x_n(t)\}$ $(n = 0, 1, 2, \cdots)$ may be considered as the non-negative measurable functions of $L(\Omega_0)$, where Ω_0 is the infinite interval $(-\infty, +\infty)$. It is clear that we have $x_n(t) \leqq x_0(t)$ $(n = 1, 2, \cdots)$ almost everywhere on Ω_0. Hence by a well-known result, the sequence $\{x_n(t)\}$ $(n = 1, 2, \cdots)$ contains a subsequence $\{x_{n_\nu}(t)\}$ $(\nu = 1, 2, \cdots)$ which converges weakly to a function $\bar{x}(t) \in L(\Omega_0)$. Since $\bar{x}(t)$ belongs to the smallest closed linear manifold which contains $\{x_n(t)\}$ $(n = 1, 2, \cdots)$, $\bar{x}(t)$ may also be considered as to belong to $L(\Omega)$, and it is again clear that the sequence $\{x_{n_\nu}(t)\}$ $(\nu = 1, 2, \cdots)$ converges weakly to $\bar{x}(t)$ as a sequence of elements of (AL).

Thus the proof of Theorem 10 and thereby the proof of Theorem 9 is completed.

REMARK 1. The proof given above is based on the representation theorem (Theorem 7), and in proving Theorem 7 we have made use of the transfinite induction (see Theorem 2). In order to avoid such a transfinite method, we shall prove the following Theorem 11. As will be shown later, Theorem 11 will give us another proof of Theorem 9. Theorem 11 is interesting in itself and will perhaps become a useful tool in the allied problems.

THEOREM 11. *Let (AL) be an abstract (L)-space (satisfying the conditions (I)–(IX)) which is not necessarily separable, and let $\{x_n\}$ $(n = 1, 2, \cdots)$ be an arbitrary sequence of points from (AL). Then there exists a separable closed linear subspace $(AL)'$ of (AL) which contains $\{x_n\}$ $(n = 1, 2, \cdots)$ and which is also closed in the sense of lattice (i.e., $y_1, y_2 \in (AL)'$ implies $y_1 \vee y_2, y_1 \wedge y_2 \in (AL)'$).*

Proof: Let Y be the set of all $y \in (AL)$ which is obtained from $\{x_n\}$ ($n = 1, 2, \cdots$) by operating a finite number of times the following operations:

$$\text{taking a sum: } y = y_1 + y_2, \tag{6.1}$$

$$\text{multiplication by a rational number: } y = \lambda y_1, \tag{6.2}$$

$$\text{taking a maximum: } y = y_1 \vee y_2, \tag{6.3}$$

$$\text{taking a minimum: } y = y_1 \wedge y_2. \tag{6.4}$$

Y is clearly a countable set, and $y_1, y_2 \in Y$ implies $y_1 + y_2$, λy_1, $y_1 \vee y_2$, and $y_1 \wedge y_2 \in Y$, where λ is a rational number. Let now $(AL)'$ be a closed cover of Y (in the strong topology of (AL)). We shall prove that this $(AL)'$ is the required one. Since it is clear that $(AL)'$ is separable and contains $\{x_n\}$ ($n = 1, 2, \cdots$), we have only to prove that $(AL)'$ is linear and is closed in the sense of lattice.

For this purpose, let $y_1, y_2 \in (AL)'$ and let λ be an arbitrary real number. Then there exist two sequences of points $\{y_{in}\}$ ($n = 1, 2, \cdots; i = 1, 2$) from Y and a sequence of rational numbers $\{\lambda_n\}$ ($n = 1, 2, \cdots$) such that $y_{in} \to y_i$ (strongly, $i = 1, 2$) and $\lambda_n \to \lambda$. Consequently, $y_{1n} + y_{2n} \to y_1 + y_2$ (strongly), $\lambda_n y_{1n} \to \lambda y_1$ (strongly), $y_{1n} \vee y_{2n} \to y_1 \vee y_2$ (strongly) and $y_{1n} \wedge y_{2n} \to y_1 \wedge y_2$ (strongly). Since $y_{1n} + y_{2n}$, $\lambda_n y_{1n}$, $y_{1n} \vee y_{2n}$ and $y_{1n} \wedge y_{2n}$ belong to Y for $n = 1, 2, \cdots$, $y_1 + y_2$, λy_1, $y_1 \vee y_2$ and $y_1 \wedge y_2$ must belong to $(AL)'$ which completes the proof of Theorem 11.

Proof of Theorem 9. By Theorem 11, there exists a separable closed linear subspace $(AL)'$ of (AL) which contains $\{x_n\}$ ($n = 0, 1, 2, \cdots$) and which is also closed in the sense of lattice. It is clear that $(AL)'$ itself is also an abstract (L)-space. Hence $(AL)'$ can be embedded isometrically and lattice-isomorphically *into* the concrete (L)-space (L) by Theorem 8. Thus $\{x_n\}$ ($n = 0, 1, 2, \cdots$) may be considered as a sequence of measurable functions $\{x_n(t)\}$ ($n = 0, 1, 2, \cdots$) of (L). If we now consider the case $x \geqq 0$, then $0 \leqq x_n(t) \leqq x_0(t)$ almost everywhere for $n = 1, 2, \cdots$. Hence, as we have observed above, there exists a subsequence $\{x_{n_\nu}\}$ ($\nu = 1, 2, \cdots$) of $\{x_n\}$ ($n = 1, 2, \cdots$) which converges weakly (as a sequence of points of (L)) to a point $\bar{x} \in (L)$. Since $\bar{x}$ belongs to a closed linear manifold which is spanned by $\{x_n\}$ ($n = 1, 2, \cdots$), $\bar{x}$ may also be considered as to belong to $(AL)'$, and it is clear that the sequence $\{x_{n_\nu}\}$ ($\nu = 1, 2, \cdots$) converges weakly to $\bar{x}$ as a sequence of points of $(AL)'$. Thus we have proved that there exists a subsequence $\{x_{n_\nu}\}$ ($\nu = 1, 2, \cdots$) of $\{x_n\}$ ($n = 1, 2, \cdots$) which converges weakly to $\bar{x} \in (AL)'$ as a sequence of points of $(AL)'$, and it is again clear that the sequence $\{x_{n_\nu}\}$ ($\nu = 1, 2, \cdots$) converges weakly to $\bar{x}$ as a sequence of points of (AL).

The rest of the proof may now be carried out exactly as in the preceding case. We have only to apply the mean ergodic theorem in Banach spaces (K. Yosida [15], S. Kakutani [7] or F. Riesz [10]).

Remark. Since the space $L(\Omega_0)$, where Ω_0 is the interval $(0, 1)$ or $(-\infty, +\infty)$,

is weakly complete, we can prove in the same manner that each concrete (L)-space $L(\Omega)$ is weakly complete. Hence, by Theorem 7,

THEOREM 12. *Every abstract (L)-space is weakly complete.*

For the case of the space of functions of bounded variation, this theorem was obtained by S. Banach and S. Mazur [2].

MATHEMATICAL INSTITUTE, OSAKA IMPERIAL UNIVERSITY

BIBLIOGRAPHY

1. S. Banach: Théorie des opérations linéaires, Warsaw, 1932.
2. S. Banach and S. Mazur: Zur Theorie der linearen Dimensionen, Studia Math., 4 (1933), 100–112.
3. G. Birkhoff: Dependent probabilities and the space (L), Proc. Nat. Acad. U. S. A., 24 (1938), 154–159.
4. H. Freudenthal: Teilweise geordnete Moduln, Proc. Acad. Amsterdam, 39 (1936), 641–651.
5. E. Hopf: Ergodentheorie, Ergebnisse der Math. und ihrer Grenzgebiete, 5, Berlin, 1937.
6. S. Kakutani: Iteration of bounded linear operations in complex Banach spaces, Proc. Imp. Acad. Japan, 14 (1938), 295–300.
7. S. Katutani: Mean ergodic theorem in abstract (L)-spaces, Proc. Imp. Acad. Japan, 15 (1939), 121–123.
8. L. Kantorovitch: Lineare halbgeordnete Räume, Recueil math., 44 (1937), 121–165.
9. O. Nikodym: Sur une généralisation des intégrales de M. Radon, Fund. Math., 15 (1930), 131–179.
10. F. Riesz: Some mean ergodic theorems, Journal of the London Math. Soc., 13 (1938), 274–278.
11. S. Saks: Theory of the integral, Warsaw, 1937.
12. M. H. Stone: Application of the theory of Boolean rings to general topology, Trans. Amer. Math. Soc., 41 (1937), 376–481.
13. H. Wallman: Lattices and topological spaces, Annals of Math., 39 (1938), 112–126.
14. F. Wecken: Unitärinvarianten selbstadjungierter Operatoren, Math. Ann., 116 (1938), 422-455.
15. K. Yosida: Mean ergodic theorem in Banach spaces, Proc. Imp. Acad. Japan, 14 (1938), 292–294.

Annals of Mathematics
Vol. 42, No. 2, April, 1941

TWO ESTIMATES CONNECTED WITH THE (α, β)-HYPOTHESIS[1]

By Peter Scherk

(Received April 22, 1940)

Small italics denote integers. The "number function" $D(x)$ of a set D of positive integers d is the number of the $d \leqq x$.

Let A, B, C be sets of positive integers a, b, c respectively and $A(x)$, $B(x)$, $C(x)$ their respective number functions. Let C be the set of all the numbers of the form a, b or $a + b$.

The so-called (α, β)-hypothesis reads: Let α and β be two positive real numbers with $\alpha + \beta < 1$. Let $A(x) \geqq \alpha x$ and $B(x) \geqq \beta x$ for $x = 1, 2, \cdots, n$. Then $C(n) \geqq (\alpha + \beta)n$.

This hypothesis has been treated among others by Landau, Besicovitch, and Schur.[2] The proofs of their theorems have the principle in common that the number function $C(n)$ is estimated more or less explicitly through integer expressions in the number functions of the two sets A and B; and it is only afterwards that estimates of these number functions are used. An analysis of Besicovitch's paper shows that Landau's and Schur's theorems can be derived from the estimate contained in his proof.[3]

The problem of constructing such integral estimates of $C(n)$ through expressions in the number functions of A and B seems interesting for two reasons: On the one hand one can hope to reach a deeper understanding of the (α, β)-hypothesis. On the other hand this problem seems more natural, since it involves no hypothesis about the sets A and B.

Until now two estimates of this kind are known besides that of Besicovitch.[3] These three estimates are sharp. In each of them the number of terms can be made arbitrarily large by increasing n.

In this paper two new estimates are presented each containing only a restricted number of terms. Like the earlier three inequalities they are valid for numbers

[1] This paper was presented to the American Mathematical Society on February 24, 1940.

[2] E. Landau: Die Goldbachsche Vermutung und der Schnirelmannsche Satz, Nachr. Ges. Wiss. Goettingen, 1930, pp. 255–276.

A. S. Besicovitch: On the density of the sum of two sequences of integers, Journal London Math. Soc. *10* (1935), pp. 246–248.

I. Schur: Ueber den Begriff der Dichte in der additiven Zahlentheorie, Sitz. Ber. Preuss. Akad. Wiss., 1936, pp. 269–297.

Also: E. Landau: Ueber einige neuere Fortschritte der additiven Zahlentheorie, Cambridge Tracts *35* (1937).

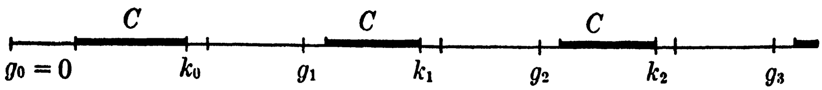
[3] P. Scherk: Bemerkungen zu einer Note von Besicovitch, Journal London Math. Soc. *14* (1939), pp. 185–192.

not lying in C and interconnected by Khintchine's "Umkehrformel" (Inversion formula).[4]

We decompose the positive numbers into intervals, such that all the integers in intervals of one kind belong to C, those in the others not: Let

$$g_0 = 0 \leqq k_0 < g_1 < k_1 < g_2 < k_2 < \cdots ;$$

$$z \subset C \quad \text{for} \quad g_i < z \leqq k_i \,; \quad z \not\subset C \quad \text{for} \quad k_i < z \leqq g_{i+1} \qquad [i = 0, 1, 2, \cdots].$$



We abbreviate $\Gamma(z) = A(z) + B(z)$.

If $k_0 \leqq n \leqq g_1$, then $C(n) \geqq \Gamma(n)$. The proof of this statement is simple.[5] Two facts follow from it: 1) The (α, β)-hypothesis is true for $k_0 < n \leqq g_1$ [and therefore up to k_1]. 2) If we suppose $\alpha + \beta \geqq 1$ instead of $\alpha + \beta < 1$, then C contains all the positive integers up to n.

If we go from the first segment $k_0 \leqq n \leqq g_1$ to any higher one $k_i \leqq n \leqq g_{i+1}$ $[i \geqq 1]$, then there exists no longer a general inequality of the form $C(n) \geqq \Gamma(n) - f(i)$, where $f(i)$ depends upon i but not on A and B.[6]

The following theorem, however, can be proved: If $k_i \leqq n \leqq g_{i+1}$, and if $C(n) < \Gamma(n)$, then there is a number $l + 1 \leqq k_i$ belonging to both A and B such that

$$C(n) \geqq \Gamma(l) + \Gamma(n - l - 1) - (i - 2).$$

This estimate contains the (α, β)-theorem for the interval $k_1 < n \leqq g_2$. For such an n satisfies at least one of the following two inequalities

$$C(n) \geqq \Gamma(n) \qquad \text{and} \qquad C(n) \geqq \Gamma(l) + \Gamma(n - l - 1) + 1.$$

The second theorem is connected with the analogous decomposition of the positive integers according to whether or not they belong to B. It can be formulated as follows: Let

$$n + 1 \not\subset C, \qquad C(n) < A(n) + B(n). \tag{1}$$

Then there is a number $m < n$ with

$$m + 1 \not\subset C, \qquad n - m \subset A,$$

---

[4] A. Khintchine: Zur additiven Zahlentheorie, Matem. Sbornik *39*, 3 (1932), pp. 27-34. One of the inequalities being symmetrical, only one other estimate is related to it.

[5] See the beginning of the proof of the first theorem. This fact was already used by Schnirelmann. Schnirelmann: Ueber additive Eigenschaften der Zahlen, Math. Annalen *107* (1933), pp. 649-690.

[6] One can readily find examples which show that $\Gamma(n) - C(n)$ can be made arbitrarily great while i remains fixed: Let

$$1 \leqq r < v, \qquad A = B = \{1, 2, \cdots, r, 2v + 1, 2v + 2, \cdots, 3v\}.$$

Hence $C = \{1, 2, \cdots, 2r, 2v + 1, 2v + 2, \cdots, 3v + r, 4v + 2, 4v + 3, \cdots, 6v\}$ $k_1 = 3v + r$ and $\Gamma(k_1) - C(k_1) = 2(v + r) - (2r + v + r) = v - r$.

such that

$$(2)\quad \begin{aligned}&(C(n) - C(n - m)) + (C(n) - C(m + 1))\\ &\qquad \geqq A(m + 1) + A(n - m - 1) + B(n) - (i - 1).\end{aligned}$$

Here $i + 1$ equals the number of numbers $r \leqq n$ with $r + 1 \not\subset B$ and either $r \subset B$ or $r = 0$.

Both theorems can be slightly refined, but they are already sharp in the given form, in the sense that there is a pair of sets A, B such that in each of the infinitely many estimates of this kind m can be chosen such that equality holds, but that for any choice of m in none of them the greater sign is valid.[7]

Throughout this paper we abbreviate $D(x, y) = D(y) - D(x)$ for any number function $D(x)$[8] and $\Gamma(x, y) = \Gamma(y) - \Gamma(x)$. Thus (2) can be written

$$(3)\quad \begin{aligned}&C(n - m, n) + C(m + 1, n)\\ &\qquad \geqq A(m + 1) + A(n - m - 1) + B(n) - (i - 1).\end{aligned}$$

LEMMA:[9] Let $k + 1 \not\subset C$, $0 \leqq g < h \leqq k$. Then

$$(4)\quad h - g \geqq A(g, h) + B(k - h, k - g).$$

PROOF: Let $a \subset A$, $b \subset B$. From $a = k + 1 - b$ would follow $k + 1 = a + b \subset C$. Therefore, the numbers a with $g < a \leqq h$ and the numbers $k + 1 - b$ with $g < k + 1 - b \leqq h$ are different from each other. Their number

$$A(g, h) + B(k - h, k - g)$$

is not greater than the number $h - g$ of all the numbers z with $g < z \leqq h$.

Upon adding the formula symmetrical to it to (4), we obtain

$$(5)\quad 2(h - g) \geqq \Gamma(g, h) + \Gamma(k - h, k - g).$$

PROOF of the first theorem:

The k_i and g_i have the same meaning as in the introduction. If $k_0 > 0$, then

$$C(k_0) = k_0 \geqq \Gamma(k_0)$$

according to (4) with $g = 0$, $h = k = k_0$. If $k_0 = 0$, this estimate is trivial. If

$$(6)\quad C(x) \geqq \Gamma(x)$$

and if $x + 1 \not\subset C$, then (6) also holds for $x + 1$ instead of x.

We may suppose (6) is not valid for all the $x \not\subset C$. Then there is a $j_1 \geqq 1$ such that (6) holds for all the $x \leqq g_{j_1}$ with $x \not\subset C$ but not for $x = k_{j_1}$. We shall show that there exists an inequality of the desired kind for all the $x \geqq k_{j_1}$ belonging to an interval $k_i \leqq x \leqq g_{i+1}$.

[7] Cf. the example given in 3.

[8] When $x < y$, $D(x, y)$ obviously indicates the number of the $d \subset D$ with $x < d \leqq y$.

[9] This lemma is a well known special case of Khintchine's "Umkehrformel"; cf. 2, 3, 4.

There is a number l_1 with $g_{j_1} \leqq l_1 < k_{j_1}$ such that

$$\Gamma(l_1) \leqq C(l_1) < C(l_1 + 1) < \Gamma(l_1 + 1). \tag{7}$$

Thus $l_1 + 1$ must lie in the intersection $A \cdot B$ of A and B. The formula (5) with

$$k = k_{j_1}, \qquad g = 0, \qquad h = k_{j_1} - l_1$$

gives

$$2C(l_1, k_{j_1}) = 2(k_{j_1} - l_1) \geqq \Gamma(k_{j_1} - l_1) + \Gamma(l_1, k_{j_1})$$

hence, because of (7)

$$2C(k_{j_1}) = 2C(l_1) + 2C(l_1, k_{j_1}) \geqq 2\Gamma(l_1) + \Gamma(k_{j_1} - l_1) + \Gamma(l_1, k_{j_1})$$
$$\geqq \Gamma(l_1) + \Gamma(k_{j_1}) + \Gamma(k_{j_1} - l_1 - 1).$$

Since $C(k_{j_1}) \leqq \Gamma(k_{j_1}) - 1$ we obtain

$$C(k_{j_1}) \geqq \Gamma(l_1) + \Gamma(k_{j_1} - l_1 - 1) + 1.$$

The numbers $k_{j_1}, k_{j_2}, \cdots, k_{j_i}$ and $l_1, l_2, \cdots, l_i$ may already have been constructed $[i \geqq 1]$ such that $l_i + 1 \subset A \cdot B$ and

$$C(x) \geqq \Gamma(l_i) + \Gamma(x - l_i - 1) - (i - 2) \tag{8}$$

for $x = k_{j_i}$. If (8) holds for x and if $x + 1 \not\subset C$, then (8) also holds for $x + 1$ instead of x on account of $l_i + 1 \subset A \cdot B$.

If (8) is not valid for all the $x \not\subset C$ with $x > k_{j_i}$, then there exists a number $j_{i+1} > j_i$ such that (8) holds for all the $x \not\subset C$ with $k_{j_i} < x \leqq g_{j_{i+1}}$ but not for $x = k_{j_{i+1}}$; and there is a number m with $g_{j_{i+1}} \leqq m < k_{j_{i+1}}$ such that

$$\begin{cases} \Gamma(l_i) + \Gamma(m - l_i - 1) - (i - 2) \leqq C(m) < C(m + 1) \\ \qquad\qquad < \Gamma(l_i) + \Gamma(m - l_i) - (i - 2). \end{cases} \tag{9}$$

Therefore

$$l_{i+1} + 1 = m - l_i \subset A \cdot B.$$

From (5) with

$$k = k_{j_{i+1}}, \qquad g = l_{i+1} + 1, \qquad h = k_{j_{i+1}} - l_i$$

we have

$$2C(m, k_{j_{i+1}}) = 2(k_{j_{i+1}} - m) = 2((k_{j_{i+1}} - l_i) - (l_{i+1} + 1))$$
$$\geqq \Gamma(l_{i+1} + 1, k_{j_{i+1}} - l_i) + \Gamma(l_i, k_{j_{i+1}} - l_{i+1} - 1)$$
$$\geqq \Gamma(l_{i+1}, k_{j_{i+1}} - l_i - 1) + \Gamma(l_i, k_{j_{i+1}} - l_{i+1} - 1) - 2.$$

On account of (9) we obtain

$$2C(k_{j_{i+1}}) = 2C(m) + 2C(m, k_{j_{i+1}}) \geqq 2(\Gamma(l_i) + \Gamma(l_{i+1}) - (i - 2)) +$$
$$+ \Gamma(l_{i+1}, k_{j_{i+1}} - l_i - 1) + \Gamma(l_i, k_{j_{i+1}} - l_{i+1} - 1) - 2$$
$$= \Gamma(l_i) + \Gamma(l_{i+1}) + \Gamma(k_{j_{i+1}} - l_i - 1) + \Gamma(k_{j_{i+1}} - l_{i+1} - 1) - 2(i - 1);$$

and since (8) should not be true for $x = k_{j_{i+1}}$

$$C(k_{j_{i+1}}) \geqq \Gamma(l_{i+1}) + \Gamma(k_{j_{i+1}} - l_{i+1} - 1) - (i - 1).$$

That accomplishes the induction, and $j_i \geqq i$ gives our theorem.[10]

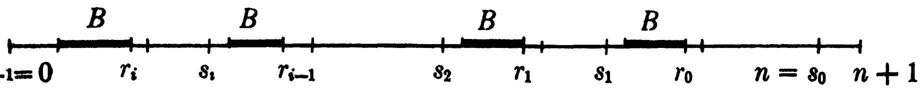
PROOF of the second theorem:

In the following a, b, c denote numbers of A, B, C respectively.

The number $n + 1$ satisfies the condition (1). Let

$$n = s_0 \geqq r_0 > s_1 > r_1 > s_2 > \cdots > r_{i-1} > s_i > r_i \geqq s_{i+1} = 0,$$

$$x \subset B \quad \text{for} \quad s_{j+1} < x \leqq r_j\,; \quad x \not\subset B \quad \text{for} \quad r_j < x \leqq s_j \quad [j = 0, 1, 2, \cdots, i].$$



From $a \leqq n - r_0$ follows $r_0 < c = a + r_0 \leqq n$; therefore

$$C(r_0, n) \geqq A(n - r_0) = A(n - r_0) + B(r_0, n).$$

Let

$$C(x, n) \geqq A(n - x) + B(x, n) \tag{10}$$

hold for $x = r_\lambda$. Then (10) is true for all the x with $s_{\lambda+1} \leqq x \leqq r_\lambda$. For (4) with

$$k = n, \qquad g = n - r_\lambda, \qquad h = n - x$$

[10] If $k_{j_1} < x < k_{j_2}$, $x \not\subset C$ then

$$C(x) \geqq \Gamma(g_{j_1}) + \Gamma(x - g_{j_1} - 1) + 1. \tag{1'}$$

Since in view of (8)

$$C(x) \geqq \Gamma(l_1) + \Gamma(x - l_1 - 1) + 1, \tag{2'}$$

it is sufficient to show that the right term of (2') is not smaller than that of (1'). Thus (1') follows from

$$2\Gamma(g_{j_1}, l_1) \geqq \Gamma(g_{j_1}, l_1) + \Gamma(x - l_1 - 1, x - g_{j_1} - 1). \tag{3'}$$

Since, on account of the lemma, the right term of (3') is not greater than $2C(g_{j_1}, l_1)$, we have only to show

$$\Gamma(g_{j_1}, l_1) \geqq C(g_{j_1}, l_1).$$

But this is evident; for, on account of (7), $C(l_1) = \Gamma(l_1)$ and $C(g_{j_1}) \geqq \Gamma(g_{j_1})$.

The following example shows that there is no estimate analogous to (2'), from which the (α, β)-hypothesis would result for the interval $k_{j_2} < x \leqq g_{j_2+1}$:

$$\begin{aligned} A &= \{1, 2, \cdots, t - 1, 2t, 2t + 1, \cdots, 3t - 1, 4t, 4t + 1, \cdots, 5t - 1\} \\ B &= \{1, 2, \cdots, t - 1, 2t, 2t + 1, \cdots, 3t - 1, 4t + 1, 4t + 2, \cdots, 5t - 1\} \end{aligned} \qquad [t \geqq 2].$$

There has not been found any example showing the same for (1').

gives

$$\begin{aligned} C(x, n) = C(x, r_\lambda) + C(r_\lambda, n) &\geqq r_\lambda - x + A(n - r_\lambda) + B(r_\lambda, n) \\ &\geqq A(n - r_\lambda, n - x) + B(x, r_\lambda) + A(n - r_\lambda) + B(r_\lambda, n) \\ &= A(n - x) + B(x, n). \end{aligned}$$

On account of (1) there exists, therefore, a number j_1 with $1 \leqq j_1 \leqq i$ such that (10) is valid for $x = r_\lambda$ [$\lambda = 0, 1, \cdots, j_1 - 1$] but not for $x = r_{j_1}$. Then (10) holds especially for $x = s_{j_1}$. Let m_1 be the greatest number x with $r_{j_1} \leqq x < s_{j_1}$ for which (10) is not true; thus

$$\begin{aligned} A(n - m_1) + B(m_1, n) > C(m_1, n) &\geqq C(m_1 + 1, n) \\ &\geqq A(n - m_1 - 1) + B(m_1 + 1, n). \end{aligned} \tag{11}$$

Since

$$B(m_1, n) = B(m_1 + 1, n) = B(r_{j_1}, n)$$

we have

$$m_1 + 1 \not\subset C, \qquad n - m_1 \subset A.^{11}$$

Furthermore

$$\text{(12)} \quad C(r_{j_1}, m_1) \geqq A(m_1 - r_{j_1}), \quad C(n - m_1 + r_{j_1}, n) \geqq A(n - m_1, n - r_{j_1}).$$

For if $r_{j_1} \neq 0$, then $r_{j_1} \subset B$;

$$a \leqq m_1 - r_{j_1} \quad \text{gives} \quad r_{j_1} < c = a + r_{j_1} \leqq m_1$$

and $n - m_1 < a \leqq n - r_{j_1}$ gives $n - m_1 + r_{j_1} < c = a + r_{j_1} \leqq n$.

Thus we obtain:

$$\begin{aligned} C(n - m_1 + r_{j_1}, n) + C(m_1 + 1, n) &\geqq A(n - m_1, n - r_{j_1}) + \\ &\quad + A(n - m_1 - 1) + B(m_1 + 1, n) \qquad \text{[see (11) and (12)]} \\ &\geqq A(n - r_{j_1}) - 1 + B(r_{j_1}, n) \\ &\geqq C(r_{j_1}, n) \qquad \text{[(10) should not hold for } x = r_{j_1}] \\ &= C(r_{j_1}, m_1) + C(m_1, n) \\ &\geqq A(m_1 - r_{j_1}) + A(n - m_1 - 1) + B(r_{j_1}, n) \qquad \text{[see (11) and (12)].} \end{aligned}$$

Let m_κ and r_{j_κ} be already defined,

$$0 < \kappa \leqq j_\kappa < i, \qquad r_{j_\kappa} \leqq m_\kappa < n,$$

$$m_\kappa + 1 \not\subset C, \qquad n - m_\kappa \subset A,$$

[11] By iterating the step leading to m_1 a descending sequence can be constructed which is essentially equivalent to Besicovitch's ascending one.

and let

$$C(n - m_\kappa + x, n) + C(m_\kappa + 1, n) \geqq A(m_\kappa - x) + A(n - m_\kappa - 1) + B(x, n) - (\kappa - 1) \tag{13}$$

hold for $x = r_{j_\kappa}$.

If (13) is true for $x = r_\lambda$, then it holds for all the x with $s_{\lambda+1} \leqq x < r_\lambda$. Since on account of $n - m_\kappa \subset A$ and $z \subset B$ for $s_{\lambda+1} < z \leqq r_\lambda$ every number $n - m_\kappa + z$ belongs to C, we have

$$C(n - m_\kappa + x, n - m_\kappa + r_\lambda) = r_\lambda - x;$$

from (13) with r_λ instead of x and (4) with

$$k = m_\kappa, \qquad g = m_\kappa - r_\lambda, \qquad h = m_\kappa - x$$

we hence derive

$$\begin{aligned} C(n - m_\kappa + x, n) &+ C(m_\kappa + 1, n) \\ &= r_\lambda - x + C(n - m_\kappa + r_\lambda, n) + C(m_\kappa + 1, n) \\ &\geqq A(m_\kappa - r_\lambda, m_\kappa - x) + B(x, r_\lambda) + A(m_\kappa - r_\lambda) + \\ &+ A(n - m_\kappa - 1) + B(r_\lambda, n) - (\kappa - 1) \\ &= A(m_\kappa - x) + A(n - m_\kappa - 1) + B(x, n) - (\kappa - 1). \end{aligned}$$

If we can choose in particular $\lambda = i$, (13) holds for $x = s_{i+1} = 0$ and, on account of $\kappa \leqq i$ and $A(m_\kappa) = A(m_\kappa + 1)$, we arrive at the assertion (3) with $m = m_\kappa$.

We suppose now that (13) is not true for $x = 0$. Then there is a number $j_{\kappa+1}$ with $j_\kappa < j_{\kappa+1} \leqq i$ such that (13) holds for $x = r_\lambda$ [$\lambda = j_\kappa, j_\kappa + 1, \cdots, j_{\kappa+1} - 1$] but not for $x = r_{j_{\kappa+1}}$. Thus we have

$$A(m_\kappa - r_{j_{\kappa+1}}) + A(n - m_\kappa - 1) + B(r_{j_{\kappa+1}}, n) - \kappa \geqq C(n - m_\kappa + r_{j_{\kappa+1}}, n) + C(m_\kappa + 1, n). \tag{14}$$

According to the above-mentioned (13) holds for $x = s_{j_{\kappa+1}}$. Let l be the greatest x with $r_{j_{\kappa+1}} \leqq x < s_{j_{\kappa+1}}$, for which (13) does not hold. Then we have

$$\begin{aligned} &A(m_\kappa - l) + A(n - m_\kappa - 1) + B(l, n) - \kappa \\ &\qquad \geqq C(n - m_\kappa + l, n) + C(m_\kappa + 1, n) \\ &\qquad \geqq C(n - m_\kappa + l + 1, n) + C(m_\kappa + 1, n) \\ &\qquad \geqq A(m_\kappa - l - 1) + A(n - m_\kappa - 1) + B(l + 1, n) - (\kappa - 1), \end{aligned} \tag{15}$$

$$B(l, n) = B(l + 1, n) = B(r_{j_{\kappa+1}}, n).$$

We put

$$m_{\kappa+1} = n - m_\kappa + l,$$

thus

$$r_{j_{\kappa+1}} \leqq l < (n - m_\kappa) + l = m_{\kappa+1} = n - (m_\kappa - l) < n,$$

and (15) can be written

$$(16)\quad \begin{cases} A(n - m_{\kappa+1}) + A(n - m_\kappa - 1) + B(r_{j_{\kappa+1}}, n) - \kappa \\ \qquad \geqq C(m_{\kappa+1}, n) + C(m_\kappa + 1, n) \\ \qquad \geqq C(m_{\kappa+1} + 1, n) + C(m_\kappa + 1, n) \\ \qquad \geqq A(n - m_{\kappa+1} - 1) + A(n - m_\kappa - 1) + B(r_{j_{\kappa+1}}, n) - (\kappa - 1). \end{cases}$$

From (16)

$$m_{\kappa+1} + 1 \not\subset C, \qquad n - m_{\kappa+1} \subset A.$$

Further

$$n - m_{\kappa+1} + r_{j_{\kappa+1}} = m_\kappa - l + r_{j_{\kappa+1}} < m_\kappa .$$

Since $r_{j_{\kappa+1}} = 0$ or $r_{j_{\kappa+1}} \subset B$ we have in analogy to (12)

$$(17)\quad \begin{cases} C(n - m_{\kappa+1} + r_{j_{\kappa+1}}, m_\kappa) \geqq A(n - m_{\kappa+1}, m_\kappa - r_{j_{\kappa+1}}) \\ C(n - m_\kappa + r_{j_{\kappa+1}}, m_{\kappa+1}) \geqq A(n - m_\kappa, m_{\kappa+1} - r_{j_{\kappa+1}}). \end{cases}$$

Therefore

$$\begin{aligned} C(m_{\kappa+1} + 1, n) &+ C(n - m_{\kappa+1} + r_{j_{\kappa+1}}, n) \\ &= C(m_{\kappa+1} + 1, n) + C(m_\kappa + 1, n) + C(n - m_{\kappa+1} + r_{j_{\kappa+1}}, m_\kappa) \\ &\geqq A(n - m_{\kappa+1} - 1) + A(n - m_\kappa - 1) + B(r_{j_{\kappa+1}}, n) - \\ &- (\kappa - 1) + A(n - m_{\kappa+1}, m_\kappa - r_{j_{\kappa+1}}) \quad \text{[see (16) and (17)]} \\ &\geqq A(m_\kappa - r_{j_{\kappa+1}}) + A(n - m_\kappa - 1) + B(r_{j_{\kappa+1}}, n) - \kappa \\ &\geqq C(n - m_\kappa + r_{j_{\kappa+1}}, n) + C(m_\kappa + 1, n) \\ &= C(n - m_\kappa + r_{j_{\kappa+1}}, m_{\kappa+1}) + C(m_{\kappa+1} + 1, n) + C(m_\kappa + 1, n) \\ &\geqq A(n - m_\kappa, m_{\kappa+1} - r_{j_{\kappa+1}}) + A(n - m_{\kappa+1} - 1) + A(n - m_\kappa - 1) + \\ &+ B(r_{j_{\kappa+1}}, n) - (\kappa - 1) \\ &\geqq A(m_{\kappa+1} - r_{j_{\kappa+1}}) + A(n - m_{\kappa+1} - 1) + B(r_{j_{\kappa+1}}, n) - \kappa. \end{aligned}$$

That finishes the induction.[12]

YALE UNIVERSITY.

[12] In analogy to footnote 10 one can prove: Let $s_{j_2} < x < s_{j_1}$, $x \subset B$; then $C(n - s_{j_1} + x, n) + C(s_{j_1}, n) \geqq A(n - s_{j_1}) + A(s_{j_1} - x) + B(x, n) + 1$. From (11) follows

$$(4') \qquad C(m_1, n) = A(n - m_1 - 1) + B(m_1, n) = A(n - m_1) + B(s_{j_1}, n) - 1.$$

From (13) with $\kappa = 1$ and (4′) follows

$$(5') \qquad C(n - m_1 + x, n) \geqq A(m_1 - x) + B(x, s_{j_1}).$$

Furthermore according to (10) with s_{j_1} instead of x:

$$(6') \qquad C(s_{j_1}, n) \geqq A(n - s_{j_1}) + B(s_{j_1}, n).$$

(4′) and (6′) together give

$$(7') \qquad C(m_1, s_{j_1}) \leqq A(n - s_{j_1}, n - m_1) - 1.$$

Finally

$$(8') \qquad \begin{cases} C(n - s_{j_1} + x, n - m_1 + x) \geqq A(n - s_{j_1}, n - m_1) & [\text{for } x \subset B] \\ \qquad\qquad \geqq C(m_1, s_{j_1}) + 1 & [\text{see } (7')] \\ \qquad\qquad \geqq A(m_1 - x, s_{j_1} - x) + 1 & [x \subset B]. \end{cases}$$

By adding (5′), (6′), and (8′) we obtain the assertion.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

SUR LE THÉORÈME DE LEBESGUE-NIKODYM

PAR JEAN DIEUDONNÉ

(Received May 16, 1940)

INTRODUCTION. Un des théorèmes les plus importants de la théorie de l'intégrale de Lebesgue est le suivant: si E désigne l'ensemble des fonctions réelles continues $x(t)$ dans l'intervalle $0 \leq t \leq 1$, toute fonctionelle linéaire positive $L(x)$ définie dans E peut se mettre, d'une manière et d'une seule, sous la forme

$$L(x) = \int_0^1 y(t)x(t)\,dt + S(x) \tag{1}$$

où $y(t)$ est une fonction mesurable positive, bien définie dans l'intervalle $[0, 1]$, à l'exception des points d'un ensemble de mesure nulle, et $S(x)$ une fonctionnelle linéaire positive *singulière*, c'est-à-dire jouissant de la propriété suivante: il existe un ensemble de mesure nulle H, contenu dans l'intervalle $[0, 1]$, tel que, pour toute fonction mesurable positive $z(t)$, nulle en tout point de H, on ait $S(z) = 0$. Ce théorème, généralisé par O. Nikodym à l'intégrale de Radon-Stieltjes[1] (où les fonctions intégrées sont définies dans un ensemble quelconque) est connu dans la littérature sous le nom de *théorème de Lebesgue-Nikodym*.

Dans un important mémoire sur les Opérations linéaires, publié récemment dans ce journal,[2] M. F. Riesz a montré l'extrême généralité de la décomposition qui apparaît dans la formule (1). En prenant comme ensemble E, non plus un ensemble de *fonctions*, mais un ensemble d'éléments de nature quelconque, possédant seulement quelques-unes des propriétés des ensembles de fonctions réelles intervenant dans le théorème de Lebesgue-Nikodym (notamment en ce qui concerne la structure d'*ordre* de cet ensemble, et sa structure de *groupe abélien*), il parvient néanmoins à montrer, par des moyens fort simples, que, si U est une fonction linéaire positive définie sur E, toute autre fonction linéaire positive L sur E peut se mettre, d'une manière et d'une seule, sous la forme $L = V + S$, où V joue le rôle de la fonctionnelle "absolument continue" $\int_0^1 y(t)x(t)\,dt$ du second membre de (1), et S celui de la partie "singulière", de la manière suivante: S est "disjointe" de U, c'est-à-dire que l'on a inf $(U, S) = 0$ (autrement dit, il n'existe pas de fonction linéaire positive, autre que 0, inférieure à la fois à U et S); et V appartient à la plus petite "famille complète" de fonctions linéaires positives contenant U (une telle famille étant un ensemble de fonctions linéaires positives caractérisé par les propriétés suivantes: il contient

[1] Voir par exemple *S. Saks, Theory of the Integral*, New York, G. E. Stechert, 1937.

[2] *F. Riesz, Sur quelques notions fondamentales dans la théorie générale des opérations linéaires*, Annals of Math., *41*, (1940), pp. 174–206.

la somme de deux quelconques de ses éléments, les minorants d'un quelconque de ses éléments, et les bornes supérieures de ses parties majorées).

Toutefois, la théorie de F. Riesz ne permet pas de donner à la partie "absolument continue" de cette décomposition une forme aussi précise que dans la formule (1); cela est dû en premier lieu au fait que le *produit* de deux éléments de l'ensemble E n'est pas défini.

Dans le même temps que je prenais connaissance du mémoire de F. Riesz, N. Bourbaki a bien voulu me communiquer un travail manuscrit[3] sur l'Intégration rédigé en 1939, indépendamment des travaux de F. Riesz, et qui contient, entre autres, une généralisation du théorème de Lebesgue-Nikodym ayant beaucoup de points communs avec les résultats de ce dernier. Chez N. Bourbaki, les éléments de l'ensemble E sont des fonctions comme dans le cas classique, mais sans les restrictions de "mesurabilité" habituelles; en fait, une grande partie de ses résultats n'utilise que les propriétés attribuées par F. Riesz à son ensemble E. En outre, tandis que les structures d'ordre n'interviennent, dans le mémoire de F. Riesz, que par les notions de borne supérieure et borne inférieure, les méthodes de N. Bourbaki reposent sur un usage prépondérant des "limites dans un ensemble ordonné filtrant";[4] le résultat fondamental auquel on aboutit ainsi (et que je reproduis avec l'autorisation de l'auteur) est le suivant: appelons *partition* d'une fonction positive $x \in E$, toute suite finie (x_i) de fonctions positives appartenant à E, et telles que $\sum_i x_i = x$; et *ordonnons* l'ensemble $\mathcal{P}(x)$ des partitions de x, en posant $(x_i) \leq (x'_j)$ si, pour toute fonction x_i de la première partition, il existe une suite finie extraite de la seconde, et qui soit une partition de x_i ; $\mathcal{P}(x)$, ainsi ordonné, est *filtrant* (à droite). Ceci posé, soit $\varphi(u_1, u_2, \cdots, u_p)$ une fonction de p variables réelles, définie dans tout l'espace numérique R^p, *lipschitzienne* (c'est-à-dire telle qu'il existe une constante c satisfaisant à l'identité $|\varphi(u_1, u_2, \cdots, u_p) - \varphi(u'_1, u'_2, \cdots, u'_p)| \leq c \cdot \mathrm{Max}\, |u_k - u'_k|$) et *positivement homogène* (c'est-à-dire que, pour tout $\lambda > 0$, $\varphi(\lambda u_1, \lambda u_2, \cdots, \lambda u_p) = \lambda \cdot \varphi(u_1, u_2, \cdots, u_p)$). Soient d'autre part $I_1, I_2, \cdots, I_p$, p fonctions linéaires "relativement bornées" sur E (différences de deux fonctions linéaires positives); alors, pour tout $x \in E$ et ≥ 0, la formule

$$(2) \qquad J(x) = \lim_{\mathcal{P}(x)} \sum_i \varphi(I_1(x_i), I_2(x_i), \cdots, I_p(x_i))$$

définit une fonction linéaire relativement bornée sur E (la limite étant prise suivant le filtre des sections de l'ensemble ordonné filtrant $\mathcal{P}(x)$); cette fonction se note $\varphi(I_1, I_2, \cdots, I_p)$.[5]

[3] Ce manuscrit n'est pas destiné à une publication immédiate, mais constitue un premier projet du fascicule des "*Eléments de Mathématique*" de cet auteur, qui sera consacré à la théorie de l'Intégration. Signalons à ce propos que nous suivons ici la terminologie de ces "Eléments" en ce qui concerne la Théorie des Ensembles et la Topologie générale.

[4] Cette notion de limite n'est autre que celle introduite par E. H. Moore et H. L. Smith ("*A general theory of limits*," Amer. Journ. of Math., *44* (1922), p. 102).

[5] F. Riesz définit aussi ces "fonctions de fonctions" linéaires, mais d'une manière tout autre et plus détournée.

Ce résultat permet à N. Bourbaki de définir une *topologie d'espace localement convexe* sur l'ensemble F des fonctions linéaires relativement bornées sur E; et, à partir de cette topologie, de donner une nouvelle définition de la "plus petite famille complète" (au sens de F. Riesz) contenant une fonction linéaire positive U. De plus, cette définition topologique donne un moyen de préciser la forme de la partie "absolument continue" de la décomposition de Riesz, en adjoignant de nouveaux éléments à E par une opération de "complétion", et en définissant le produit d'un de ces éléments (qui ne sont plus des fonctions, contrairement aux éléments de E), et d'un élément de E; on montre alors que la partie "absolument continue" de la décomposition de Riesz peut s'écrire $U(yx)$, où y est un des éléments de l'ensemble E "complété" (U étant convenablement "prolongée" à cet ensemble); on arrive ainsi à une généralisation parfaite du théorème de Lebesgue-Nikodym.[6]

Toutefois, cette dernière partie du travail de N. Bourbaki (dont le point capital est la démonstration de l'identité de la partie de F définie par voie topologique, et de la "plus petite famille complète" correspondante) suppose essentiellement que les éléments de E sont des *fonctions*. Il restait à examiner la possibilité d'arriver à des résultats analogues en demeurant dans la voie suivie par F. Riesz, c'est-à-dire sans supposer le caractère fonctionnel des éléments de l'ensemble E; c'est ce que je fais dans ce qui suit. J'y utilise les méthodes topologiques de N. Bourbaki, ainsi que la formule (2) (qui ne suppose pas que les éléments de E sont des fonctions);[7] quant au raccord avec la théorie de F. Riesz, qui reste le point délicat de la démonstration, il se fait en adaptant au cas "abstrait" une idée de J. von Neumann, appliquée par ce dernier à la démonstration du théorème de Lebesgue-Nikodym classique, et qui consiste à passer par l'intermédiaire de l'espace des "fonctions de carré sommable."[8]

1. Nous prenons comme point de départ un ensemble E sur lequel est défini, d'une part une structure d'*ordre*, d'autre part une structure d'*espace vectoriel* par rapport au corps des nombres réels; nous supposons en outre que les conditions suivantes sont remplies:

(I) E est *réticulé* ("lattice"), autrement dit, quels que soient x et y dans E, il existe les éléments $\inf(x, y)$ et $\sup(x, y)$.

(II) La relation $x \leq y$ entraîne $x + z \leq y + z$ quel que soit z.

(III) Les conditions $x \geq 0$, $\lambda \geq 0$ (λ réel) entraînent $\lambda x \geq 0$. On pose $|x| = \sup(x, -x)$; on démontre sans peine les identités $|x + y| \leq |x| + |y|$, $|\lambda x| = |\lambda|\cdot|x|$; si $x^+ = \sup(x, 0)$, $x^- = \sup(-x, 0)$ on a $x = x^+ - x^-$, $|x| = x^+ + x^-$. L'ensemble des éléments ≥ 0 de E sera désigné par E_+; il

[6] En ce qui concerne les travaux de N. Bourbaki sur l'Intégration on pourra consulter un article de A. Weil sur la théorie des probabilités (*Revue Scientifique* (Revue rose), 1940.

[7] A l'exception de cette formule, tous les résultats de N. Bourbaki utilisés dans ce travail sont donnés avec leur démonstration.

[8] Voir par exemple, *J. von Neumann, On rings of Operators, III*, Annals of Math., *41* (1940), p. 127–129.

satisfait aux conditions imposées par F. Riesz à son "domaine fondamental" (loc. cit., §I).

Une fonction réelle U sur E est *linéaire* si $U(x + y) = U(x) + U(y)$, et $U(\lambda x) = \lambda U(x)$; elle est *positive* si $U(x) \geq 0$ quel que soit $x \in E_+$; elle est *relativement bornée* si, quel que soit $y \in E_+$, $| U(x) |$ est borné pour l'ensemble des $x \in E$ tels que $| x | \leq y$. Toute fonction relativement bornée est la différence de deux fonctions linéaires positives (N. Bourbaki, loc. cit.).

On désignera par F l'ensemble des fonctions linéaires relativement bornées; si on écrit $U \geq V$ lorsque $U - V \geq 0$, F satisfait aux conditions (I), (II) et (III). D'après la formule (2), on a pour tout $x \in E_+$ et tout $U \in F$, $| U |(x) = \sup \sum_i | U(x_i) |$ pour toutes les *partitions finies* (x_i) de x. On désignera par F_+ l'ensemble des fonctions linéaires positives.

On peut, avec N. Bourbaki, définir sur F une structure d'espace vectoriel *localement convexe*, à l'aide de la famille de *pseudo-normes* $N_x(X) = | X |(x)$ (x fixe dans E_+); F, muni de cette structure, est *complet*. En effet, soit $\mathfrak{F}$ un filtre de Cauchy sur F; quel que soit $x \in E_+$, il existe un ensemble $A \in \mathfrak{F}$ tel que, pour X et Y quelconques dans A, $| X - Y |(x) \leq \epsilon$, et à fortiori $| X(x) - Y(x) | \leq \epsilon$. Si $A(x)$ désigne l'ensemble des nombres $X(x)$ pour $X \in A$, et $\mathfrak{F}(x)$ le filtre formé par les ensembles $A(x)$ sur la droite numérique R, $\mathfrak{F}(x)$ est un filtre de Cauchy, qui converge donc vers une limite, qu'on notera $X_0(x)$. Montrons que X_0 est la limite de $\mathfrak{F}$ dans F; pour tout $X \in A$, et toute partition finie (x_i) de x, on a

$$\sum_i | X_0(x_i) - X(x_i) | = \sum_i \lim_{\mathfrak{F}(x_i)} | Y(x_i) - X(x_i) | \leq \sup_{Y \in A} \sum_i | Y(x_i) - X(x_i) |$$

$$\leq \sup_{Y \in A} | Y - X | (x) \leq \epsilon$$

ce qui montre d'abord que $X_0 - X \in F$, d'où $X_0 \in F$, puis que $| X_0 - X | (x) \leq \epsilon$, ce qui achève la démonstration.

Dans F, tout ensemble *majoré* A possède une borne supérieure (F. Riesz, loc. cit., §II), et toute ensemble minoré une borne inférieure; si en outre A est ordonné *filtrant à droite*, sa borne supérieure est aussi la limite (dans la topologie définie ci-dessus) du *filtre des sections* sur A; on peut d'ailleurs retrouver directement cette propriété en partant du fait que F est complet, et en déduire ensuite le résultat de F. Riesz. Si X_0 est la limite de l'ordonné filtrant A, on a $X_0(x) = \lim_{X \in A} X(x)$, quel que soit $x \in E$.

Remarquons encore que, dans F, F_+ est un ensemble *fermé*, et que les fonctions X^+, X^-, $| X |$ sont *uniformément continues*.

2. Nous supposerons à partir de maintenant que E est muni, en outre des structures précédentes, d'une structure d'*anneau commutatif*, c'est-à-dire que le *produit* xy de deux éléments de E est toujours défini, est associatif, commu-

tatif, distributif par rapport à l'addition et del que $\lambda(xy) = (\lambda x)y$ pour λ réel; enfin, nous faisons l'hypothèse suivante:

(IV) Quels que soient y, z dans E, la relation $x \geq 0$ entraîne

$$\sup(xy, xz) = x \cdot \sup(y, z)$$

Indiquons d'abord quelques conséquences de cette hypothèse. En premier lieu, si $y \leq z$, on a $xy \leq xz$ si $x \geq 0$; en particulier, si $x \geq 0$, $y \geq 0$, $xy \geq 0$; si $x \geq 0$ ou $x \leq 0$, on a $x^2 = (-x)^2 \geq 0$. Si $x \geq 0$ et y est quelconque, $|xy| = x \cdot |y|$; si donc x et y sont quelconques et qu'on pose $x' = |x|$, $y' = |y|$, on aura $|xy| = |x^+y - x^-y| \geqq |(|x^+y| - |x^-y|)| = |(x^+ - x^-)y'| = |xy'| = x'y'$, et $|xy| \leqq |x^+y| + |x^-y| = (x^+ + x^-)y' = x'y'$, donc $|xy| = x'y' = |x| \cdot |y|$. De plus $xx' = x'^2$; or $xx' = (x^+)^2 - (x^-)^2$, donc $xx' \leqq (x^+)^2 + (x^-)^2$; mais $x'^2 = (x^+)^2 + (x^-)^2 + 2x^+x^-$, ce qui montre que $x^+x^- = 0$; on en tire que, pour x quelconque, $x^2 = (x^+)^2 + (x^-)^2 \geq 0$.

Soit alors $X \in F_+$; quels que soient λ et μ réels, et x et y quelconques dans E, on a $X((\lambda x + \mu y)^2) \geq 0$, d'où aussitôt l'inégalité de Schwartz

$$|X(xy)| \leq (X(x^2))^{\frac{1}{2}}(X(y^2))^{\frac{1}{2}} \tag{3}$$

et, comme conséquence immédiate, l'inégalité de Minkowski

$$(X((x + y)^2))^{\frac{1}{2}} \leq (X(x^2))^{\frac{1}{2}} + (X(y^2))^{\frac{1}{2}} \tag{4}$$

3. Considérons à présent une fonction linéaire positive U sur E, sur laquelle nous ferons d'abord les hypothèses suivantes:

(V) La relation $U(|x|) = 0$ entraîne $x = 0$.

(VI) Quel que soit $y \in E$, il existe un nombre fini $\|y\| \geq 0$ tel que, pour tout $x \in E$,

$$|U(yx)| \leq \|y\| \cdot U(|x|) \tag{5}$$

La première condition exprime que $U(|x|)$ est une *norme* sur E; la seconde, que, si on désigne par U_y la fonction linéaire $x \to U(yx)$, U_y est *continue* lorsqu'on topologise E par la norme $U(|x|)$.

Pour énoncer notre dernière hypothèse, désignons par S l'ensemble des $y \in E_+$ tels que, pour *tout* $x \in E_+$, on ait $yx \leq x$. Notre hypothèse est la suivante:

(VII) Quels que soient $X \in F$ tel que $0 \leq X \leq U$, et $x \in E$, on a

$$X(|x|) = \sup_{|y| \in S} |X(yx)| \tag{6}$$

On en déduit que, si $U(xy) = 0$ quel que soit $y \geq 0$, on a $x = 0$; en effet, on a alors, pour tout $y \in E$, $|U(xy)| \leq |U(xy^+)| + |U(xy^-)| = 0$, d'où $U(xy) = 0$; d'après (6) on en tire $U(|x|) = 0$, et cela entraîne $x = 0$ d'après (V). Il en résulte en particulier que, si $x^2 = 0$, on a $x = 0$, en vertu de l'inégalité de Schwartz.

4. Soit F_U l'ensemble des U_y, lorsque y parcourt E; d'après ce qui précède, l'application $y \rightarrow U_y$ est une application *biunivoque* de E sur F_U. Elle est évidemment *linéaire*; en outre, c'est un *isomorphisme* de la structure d'*ordre* de E sur celle de F_U. En effet, il est clair que $y' \leq y''$ entraîne $U_{y'} \leq U_{y''}$; mais réciproquement, $U_{y'} \leq U_{y''}$ entraîne $y' \leq y''$: il suffit de montrer que $U(yx) \geq 0$ quel que soit $x \geq 0$ entraîne $y \geq 0$. Or cette hypothèse entraîne en particulier $U(yy^-) \geq 0$, c'est-à-dire $U(-(y^-)^2) \geq 0$; comme $(y^-)^2 \geq 0$, ce n'est possible que si $U((y^-)^2) = 0$, d'où $(y^-)^2 = 0$, et par suite $y^- = 0$, c'est-à-dire $y \geq 0$.

A la topologie de F_U correspond sur E celle définie par les pseudo-normes $N'_y(x) = U_y(|x|)$ pour $y \geq 0$; cela résulte de l'identité $|U_z| = U_{|z|}$, que nous allons démontrer en nous appuyant sur l'hypothèse (VII), la formule (2), et un raisonnement emprunté au mémoire de N. Bourbaki. Il suffit évidemment de voir que, pour tout $x \geq 0$, $U(|z|x) \leq \sup \sum_i |U(zx_i)|$ pour toutes les partitions (x_i) de x. Or, d'après (VII), on a $U(|z|x) = \sup_{|y| \in S} |U(zxy)|$; mais

$$|U(zxy)| \leq |U(zxy^+)| + |U(zxy^-)| \leq \sum_i |U(zx_i)|$$

où (x_i) est la partition de x formée des trois éléments $x_1 = xy^+$, $x_2 = xy^-$, $x_3 = x - x|y|$; d'où la proposition.

L'application $y \rightarrow U_y$ se prolonge donc par continuité, en une application biunivoque du *complété* E_U de E, muni de la structure uniforme définie par les pseudo-normes N'_y, sur l'*adhérence* $\bar{F}_U$ de F_U dans F. Pour tout $y \in E_U$, on désignera encore par U_y l'élément de $\bar{F}_U$ qui lui correspond; on *ordonnera* E_U en y transportant la structure d'ordre de $\bar{F}_U$; cette structure transportée prolonge celle de E, d'après ce qui précède; l'ensemble E_U^+ des $y \geq 0$ est l'*adhérence* de E_+ dans E_U. En effet, il suffit de voir que, si G_U désigne l'ensemble des U_y tels que $y \in E_+$, $\bar{G}_U = \bar{F}_U \cap F_+$; or, l'ensemble $\bar{F}_U \cap F_+$, étant fermé, contient $\bar{G}_U$; et d'autre part, si $X \geq 0$ appartient à $\bar{F}_U$, à tout $x \in E_+$ et tout $\epsilon > 0$ correspond un $y \in E$ tel que $|X - U_y|(x) \leq \epsilon$, d'où, puisque $|X| = X$, et $|U_y| = U_{|y|}$, $|X - U_{|y|}|(x) \leq \epsilon$, ce qui montre que $X \in \bar{G}_U$. On en conclut aisément que, si $X \in \bar{F}_U$, $|X| \in \bar{G}_U$, d'après la continuité de la fonction $|X|$ dans F.

Il est clair par ailleurs que la structure d'espace vectoriel de E se prolonge à E_U, en une structure isomorphe à celle de $\bar{F}_U$; cette structure et la structure d'ordre de E_U satisfont aux axiomes (I), (II) et (III). En outre, *tout ensemble majoré dans E_U admet une borne supérieure*, car la borne supérieure, dans F, d'une partie majorée de $\bar{F}_U$ appartient à $\bar{F}_U$: c'est immédiat pour une partie *finie* $\left(\text{d'après la formule } \sup(X, Y) = \dfrac{X + Y + |X - Y|}{2}\right)$, et il en résulte que la borne supérieure d'une partie majorée quelconque de $\bar{F}_U$ est la limite d'un ensemble filtrant à droite sur $\bar{F}_U$, donc appartient à $\bar{F}_U$ (qui est fermé).

5. Pour une valeur *fixe* de $x \in E$, l'application linéaire $y \rightarrow yx$ de E dans E est *continue*; en effet, d'après la condition (VI), on a, pour tout $z \in E_+$, $N'_z(yx) = U(|yx|z) \leq ||x|| \cdot U(|y|z) = ||x|| \cdot N'_z(y)$.

On peut donc *prolonger* par continuité cette fonction à E_U ; on a ainsi défini le produit yx pour tout $y \in E_U$ et tout $x \in E$ (le produit yy', pour deux éléments de E_U, n'est pas défini en général!) Si x et x' sont deux éléments quelconques de E, on a $(yx)x' = y(xx')$. En effet, les deux membres sont des fonctions continues de y, égales dans E, qui est partout dense dans E_U, donc elles sont encore égales dans E_U. On montre de même les propriétés de distributivité $(y + y')x = yx + y'x$, $y(x + x') = yx + yx'$, et que, pour λ réel, $(\lambda y)x = y(\lambda x) = \lambda(yx)$. Si on remarque que la fonction sup (y, z) est uniformément continue dans $E_U \times E_U$, on démontre de la même manière que sup $(y, y')\cdot x =$ sup $(yx, y'x)$, et que $y\cdot$sup $(x, x') =$ sup (yx, yx') en prolongeant l'identité (IV).

Enfin, les fonctions $U(yx)$ et $U_y(x)$, pour x fixe dans E, sont linéaires en y, continues dans E et identiques; elles sont donc encore identiques pour tout $y \in E_U$.

6. Nous sommes maintenant en mesure d'énoncer le résultat final auquel nous voulons parvenir. Désignons par K ce que F. Riesz appelle "la plus petite famille complète contenant U"; cette famille est une partie de F_+ qui s'obtient de la manière suivante (F. Riesz, loc. cit., §V): on considère l'ensemble K' des fonctions linéaires positives X telles que $X \leq \lambda U$ pour une valeur convenable de $\lambda > 0$; K est l'ensemble des bornes supérieures de toutes les parties majorées de K'. Nous allons montrer que *l'ensemble K est identique à l'ensemble $\bar{G}_U$*. Toute fonction $X \in F_+$ se mettant, d'après le théorème fondamental de F. Riesz (loc. cit., th. 14) sous la forme $Y + Z$, où $Y \in K$ et où Z est *disjointe* de U, on aura $Y = U_y$ avec $y \in E_U$, d'après ce qui précède, et on obtiendra ainsi l'énoncé généralisant entièrement le théorème de Lebesgue-Nikodym aux fonctions linéaires abstraites.

Nous allons commencer par montrer que $K \subset \bar{G}_U$; comme toute partie majorée de $\bar{G}_U$ admet une borne supérieure appartenant à $\bar{G}_U$, il nous suffira de voir que $K' \subset \bar{G}_U$. Autrement dit, nous allons démontrer que, *si* $0 \leq X \leq \lambda U$ $(\lambda > 0)$, *il existe un $y \in E_U$ tel que $X = U_y$*.

Remarquons pour cela que, d'après (V) et les inégalités (3) et (4), $(U(x^2))^{\frac{1}{2}}$ est une *norme* dans E, et que, d'après (3), la fonction $(y, x) = U(yx)$ possède les propriétés d'un *produit scalaire*. Il en résulte aussitôt que, si on *complète* E, topologisé par la norme $(U(x^2))^{\frac{1}{2}}$, on obtient un *espace de Hilbert*[9] H. Comme, d'après (3), on a, pour tout $y \geq 0$,

$$N'_y(x) \leq (U(y^2))^{\frac{1}{2}}(U(x^2))^{\frac{1}{2}}$$

la topologie définie sur E par la norme $(U(x^2))^{\frac{1}{2}}$ est *plus fine* que celle définie par les pseudo-normes $N'_y(x)$; il s'ensuit que l'application identique φ de E, considéré comme sous-espace de H, sur E, considéré comme sous-espace de E_U,

[9] Par "espace de Hilbert," nous entendons un espace satisfaisant aux axiomes A, B, E énoncés au chap. I du mémoire de J. von Neumann, "*Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren*," Math. Ann., *102* (1930), p. 49. Aucune hypothèse n'est faite sur le nombre de dimensions de l'espace, qui peut être, éventuellement, fini ou supérieur au dénombrable.

se prolonge en une application linéaire *continue* de H dans E_U, application que nous désignerons encore par φ. Remarquons maintenant que *toute* forme linéaire continue sur H est un produit scalaire (y, x), avec y fixe dans H. Or, pour tout x fixe dans E, (y, x) est une fonction continue de y dans H, qui coincide avec $U(yx) = U(\varphi(y)x)$ lorsque $y \in E$; mais $U(yx)$, considérée comme fonction de y, est continue dans E_U, donc $U(\varphi(y)x)$ est continue dans H. Par suite, (y, x) et $U(\varphi(y)x)$, qui coïncident dans la partie partout dense E de H, coïncident dans H.

Ceci posé, l'hypothèse $X \leq \lambda U$ entraîne, pour tout $y \in E$, $|X(yx)| \leq \lambda U(|y|\cdot|x|) = \lambda N'_{|y|}(x)$, autrement dit, l'application $x \rightarrow X(yx)$, que nous désignerons par X_y, est *continue* dans E, muni de la structure définie par les pseudo-normes N'_y ; à fortiori, elle est continue dans E, considéré comme sous-espace de l'espace de Hilbert H. Elle se prolonge donc dans H, et par suite, il existe $z_y \in E_U$ tel que $X_y = U_{z_y}$.

Il est immédiat que l'application $y \rightarrow z_y$ de E dans E_U est linéaire. Si $y \geq 0$, on a $z_y \geq 0$, car alors $U_{z_y}(x) = X_y(x) \geq 0$ pour tout $x \geq 0$. Enfin, z_y est fonction *continue* de y; en effet, pour tout $x \in E_+$ on a $N'_x(z_y) = U(|z_y|x) = U_{|z_y|}(x) = |U_{z_y}|(x) = |X_y|(x) = X_{|y|}(x) \leq \lambda N'_x(y)$ d'après l'identité démontrée au §4.

Faisons alors décrire à y l'ensemble S (§3); d'après la condition (IV), cet ensemble ordonné est filtrant à droite; il est *majoré* dans E_U, car $U_y \leq U$ quel que soit $y \in S$; son filtre des sections $\mathfrak{F}$ a donc une limite e dans E_U. En vertu de la continuité de z_y, cette fonction tend vers une limite z_e suivant le filtre $\mathfrak{F}$; pour tout x fixe dans E_+, d'après la continuité de $U(yx)$ dans E_U, on a donc $U(z_e x) = \lim_{\mathfrak{F}} U(z_y x) = \lim_{\mathfrak{F}} X(yx)$; mais, d'après (VII), $\lim_{\mathfrak{F}} X(yx) = X(x)$, ce qui démontre la première partie du théorème.

7. Il nous reste à voir que $\bar{G}_U \subset K$; d'après le théorème de décomposition de F. Riesz, tout $X \in \bar{G}_U$ peut se mettre d'une manière et d'une seule sous la forme $Y + Z$, où $Y \in K$ et où Z est disjointe de U; comme $K \subset \bar{G}_U$, $Z = X - Y$ appartient à $\bar{F}_U$, et comme $Z \geq 0$, $Z \in \bar{G}_U$; la proposition sera démontrée si on fait voir que toute fonction $Z \in \bar{G}_U$, disjointe de U, est nécessairement 0.

Nous utiliserons pour cela le critère suivant: si une fonction X appartient à $\bar{G}_U$, à tout $y \in E_+$ et tout $\epsilon > 0$ correspond un nombre $\eta > 0$ tel que les conditions $x \in E_+$, $x \leq y$, $U(x) \leq \eta$ entraînent $|X(x)| \leq \epsilon$ (critère d'"absolue continuité"; nous en empruntons l'énoncé et la démonstration au travail de N. Bourbaki). En effet, il existe par hypothèse $z \in E$ tel que $|X - U_z|(y) \leq \delta$, d'où, pour la partition $(x, y - x)$ de y, $|X(x) - U(zx)| + |X(y - x) - U(z(y - x))| \leq \delta$ et a fortiori $|X(x) - U(zx)| \leq \delta$; par suite $|X(x)| \leq \delta + \|z\|\cdot U(x)$, ce qui est aussi petit qu'on veut si δ, puis $U(x)$, sont pris assez petits.

Si Z appartient à $\bar{G}_U$ et est disjointe de U, on a par définition inf $(Z, U) = 0$; et, pour tout $x \in E_+$ et tout $\eta \geq 0$, il existe une partition (y, z) de x telle que $U(y) + Z(z) \leq \eta$,[10] d'où en particulier $U(y) \leq \eta$; si $\eta \leq \epsilon$ est choisi de sorte

[10] *F. Riesz*, loc. cit., p. 183.

que les conditions $y \leq x$, $U(y) \leq \eta$ entraînent $Z(y) \leq \epsilon$, on aura donc $Z(x) = Z(y) + Z(z) \leq 2\epsilon$, et comme ϵ est arbitraire, $Z(x) = 0$, d'où la proposition, qui achève de démontrer le théorème.

Le même raisonnement montre aussi que la condition donnée ci-dessus pour que $X \in \bar{G}_U$, est non seulement *nécessaire*, mais aussi *suffisante*.

8. Si les éléments de E sont des *fonctions réelles finies*, définies sur un ensemble A, les axiomes (II) et (III) sont vérifiés d'eux-mêmes, et (I) entraîne (IV) (on suppose bien entendu que l'addition et la multiplication sont l'addition et la multiplication ordinaires des fonctions réelles). L'axiome (VI) est vérifié pour *toute* fonction linéaire positive U si les éléments de E sont des fonctions *bornées* sur A.

Quant à la condition (VII), N. Bourbaki a montré, dans le travail auquel nous nous sommes déjà plusieurs fois référé, qu'elle est vérifiée dans chacun des deux cas suivants: 1° E contient la fonction égale à la constante $+1$; 2° si $x \in E_+$, $\sqrt{x} \in E$. Il est vraisemblable que ces deux cas ne sont pas les seuls où la condition (VII) est réalisée; mais nous allons voir par contre, sur un exemple, que cette condition n'est pas une conséquence des autres, et que, lorsqu'elle n'est pas vérifiée, le théorème de Lebesgue-Nikodym peut se trouver en défaut.

Prenons pour A l'intervalle $0 \leq t \leq 1$, et pour E l'ensemble des fonctions réelles *continues* dans A, *nulles* au point $t = 0$, et *dérivables* en ce point. Il est immédiat que E est un espace vectoriel, et qu'il satisfait aux conditions (I), (II) et (III); en outre le produit de deux fonctions de E appartient à E, et la condition (IV) est évidemment vérifiée. Les fonctions de E sont bornées, donc (VI) est satisfaite pour toute fonction linéaire positive.

Prenons alors $U(x) = x'(0) + \int_0^1 x(t)\,dt$; il est clair que (V) est vérifiée. Par contre, nous allons voir que (VII) ne l'est pas. En effet, pour tout $y \in E$, on a $U(yx) = \int_0^1 y(t)x(t)\,dt$; si $|y(t)| \leq 1$ quel que soit $t \in A$, on a $|U(yx)| \leq \int_0^1 |x(t)|\,dt$, donc, pour tout $x \geq 0$, $|U(x) - U(yx)| \geq x'(0)$, et si $x'(0) \neq 0$, (VII) ne peut être satisfaite.

Pour montrer que, dans ce cas, le théorème de Lebesgue-Nikodym, tel que nous l'avons énoncé, n'est pas vrai, nous établirons que *U n'appartient pas à* $\bar{G}_U$. Sinon, en effet, pour tout $x \in E_+$ et tout $\epsilon > 0$, il existerait $y \in E_+$ tel que $|U - U_y|(x) \leq \epsilon$, c'est-à-dire, pour toute partition (x_i) de x, $\sum_i |U(x_i) - U(yx_i)| \leq \epsilon$, et en particulier, $|x_i'(0) + \int_0^1 (1 - y(t))x_i(t)\,dt| \leq \epsilon$ pour tout indice i. Or, on peut choisir la partition (x_i) de sorte que, pour un indice k, $x_k'(0) = x'(0)$, et $|\int_0^1 (1 - y(t))x_k(t)\,dt| \leq \epsilon$. On aurait donc $|x'(0)| \leq 2\epsilon$, quel que soit $\epsilon > 0$, ce qui est absurde, puisque x a été prise arbitraire dans E.

NANCY, FRANCE.

ANNALS OF MATHEMATICS
Vol. 42, No. 2, April, 1941

ON THE MODULAR CHARACTERS OF GROUPS

BY R. BRAUER AND C. NESBITT*

(Received, December 15, 1939)

* Presented to the American Mathematical Society, October 30, 1937.

I. INTRODUCTION[1]

1. Ordinary Representations. Group ring. The representations of a group $\mathfrak{G}$ of finite order g were first treated by *Frobenius*[2] in his theory of group characters. The coefficients of the linear transformations are taken as complex numbers, but it does not make any difference if we take them as the elements of an algebraically closed field K of characteristic 0. The theory has been extended by *I. Schur*[3] to the case where K is any field of characteristic 0. It does not mean an essential restriction, if we take K as an algebraic number field.

Instead of considering representations of $\mathfrak{G}$, we may consider representations of the group ring[4] Γ of $\mathfrak{G}$ with regard to K. This Γ is an associative algebra consisting of all symbols

$$\alpha = a_1G_1 + a_2G_2 + \cdots + a_gG_g \tag{1}$$

where $G_1, G_2, \cdots, G_g$ are the elements of $\mathfrak{G}$, and $a_1, a_2, \cdots, a_g$ are arbitrary elements of K. The equality of two such elements, their addition, and their multiplication are defined in a natural manner. The study of the representations of Γ is closely tied up with the investigation of the algebra Γ.

2. Arithmetical questions. We may also study Γ from an arithmetical point of view. Taking K as an algebraic number field, we obtain a domain of integrity $\mathfrak{J}$ if we take the a_i in (1) from the domain $\mathfrak{o}$ of the integers of K. The question arises in what manner does a prime ideal $\mathfrak{p}$ behave when considered as an ideal of $\mathfrak{J}$. The behavior of $\mathfrak{p}$ in $\mathfrak{J}$ is characterized by the structure of the residue class ring $\mathfrak{J}/\mathfrak{p}$. This ring can be considered as an algebra $\bar{\Gamma}$ over the residue class field $\bar{K} = \mathfrak{o}/\mathfrak{p}$ of the integers of K taken (mod $\mathfrak{p}$). Obviously, $\bar{\Gamma}$ is the group ring of $\mathfrak{G}$ with regard to the finite ground field $\bar{K}$. The study of the structure of $\bar{\Gamma}$ then amounts essentially to the same thing as the study of the representations of $\bar{\Gamma}$ or of $\mathfrak{G}$ by matrices with coefficients in the finite field $\bar{K}$. We thus are led to the problem of extending Frobenius' theory to the case of a modular field of reference (i.e. a field of a characteristic $p \neq 0$).

3. Modular representations. Modular representations of a group $\mathfrak{G}$ (i.e. representations of $\mathfrak{G}$ by matrices with coefficients in a modular field) were first

---

[1] In §§4–10 of the introduction, we give a short account of the theory of modular representations of a group as developed in our paper: On the modular representations of groups of finite order, University of Toronto Studies, Math. Series No. 4, 1937 (we refer to this paper as M.R.). We tried to make it unnecessary for a reader, who is familiar with the theory of representations in general, to read our former paper. An exception is formed perhaps by the proof of formula (5) below, but literature for other proofs of this formula are mentioned in footnote 10.

[2] For Frobenius' theory, see the accounts in L. E. Dickson, *Modern Algebraic Theories*, Chicago, 1926, chapter XIV; G. A. Miller, H. F. Blichfeldt, L. E. Dickson, *Theory and Application of Finite Groups*, New York 1916, chapter XIII, H. F. Blichfeldt, *Finite Collineation Groups*, Chicago 1917, chapter VI.

[3] I. Schur, Sitzungsber. Preuss. Akad., 1906, p. 164.

[4] Cf., for instance, H. Weyl, *The Classical Groups*, Princeton 1939, Chapter III.

studied by Dickson.[5] He proved that Frobenius' theory remains valid, if the characteristic p of the field is prime to the order g of $\mathfrak{G}$. Since the discriminant of Γ is a power of g, this corresponds to the case that the prime ideal $\mathfrak{p}$ in §2 is not a discriminant divisor. If, however, p divides g, then we must expect results which differ from those of Frobenius. This was shown first by a theorem of Dickson[6] concerning the splitting of the regular representation (cf. §8 below). A coherent theory of the modular representations was given by the authors in a previous paper.[7] In the following §§4–9, we shall discuss briefly our former results. We prefer, in most of what follows, to use the language of the theory of representations (instead of that of the theory of algebras or of the theory of ideals).

4. Decomposition numbers. We choose the algebraic number field K such that the absolutely irreducible representations of $\mathfrak{G}$ in the sense of Frobenius can be written with coefficients in K. Let $Z_1, Z_2, \cdots, Z_n$ be the essentially different ones among these representations, and let z_i denote the degree of Z_i. Then n is the number of classes of conjugate elements $\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_n$ in $\mathfrak{G}$.

Let p be a fixed rational prime number, and $\mathfrak{p}$ be a fixed prime ideal divisor of p in K. We may assume that the coefficients of all the Z_i are $\mathfrak{p}$-integers (i.e. numbers of the form α/β where α and β are integers of K, and β is prime to $\mathfrak{p}$). Let $\mathfrak{o}_\mathfrak{p}$ be the ring of the $\mathfrak{p}$-integers of K, and $\bar{K}$ the residue class field of $\mathfrak{o}_\mathfrak{p}$ (mod $\mathfrak{p}$) which is identical with the field $\mathfrak{o}/\mathfrak{p}$ in §2. We denote generally the residue class of an element z of K (mod $\mathfrak{p}$) by $\bar{z}$. Similarly, replacing every coefficient z in a representation Z of $\mathfrak{G}$ with coefficients in $\mathfrak{o}_\mathfrak{p}$ by its residue class $\bar{z}$, we obtain a modular representation $\bar{Z}$ with coefficients in $\bar{K}$. In this manner we may form $\bar{Z}_1, \bar{Z}_2, \cdots, \bar{Z}_n$. These modular representations will, in general, be reducible and will then split into irreducible modular representations F_κ with coefficients in $\bar{K}$. We indicate by

$$\bar{Z}_i \leftrightarrow \sum_\kappa d_{i\kappa} F_\kappa \tag{2}$$

that F_κ will appear in $\bar{Z}_i$ with some multiplicity $d_{i\kappa}$. These rational integers $d_{i\kappa} \geqq 0$, and are called the "decomposition numbers" of $\mathfrak{G}$. In the sense of §2, they describe a connection between the simple invariant subalgebras of Γ, and the prime ideal divisors of $\mathfrak{p}$ in $\mathfrak{J}$.

5. Cartan invariants. Of special importance is the regular representation $\bar{R}$ of $\mathfrak{G}$ (or $\bar{\Gamma}$) formed with regard to $\bar{K}$ as ground field. Since the group ring is no longer semisimple in the modular case, the theorem of the full reducibility of $\bar{R}$ does not hold any more. Let $U_1, U_2, \cdots, U_k$ be the distinct indecom-

[5] L. E. Dickson, Transact. Am. Math. Soc. 8, 1907, p. 389.

[6] L. E. Dickson, Bull. Amer. Math. Soc. 13, 1907, p. 477.

[7] For the following, cf. M.R., and also R. Brauer, Nat. Ac. of Sciences 25, 1939, p. 252. We refer to this last paper as R.A.

posable constituents of $\bar{R}$. Each U_κ can still be broken up into its irreducible constituents in $\bar{K}$. This splitting is of the form

$$(3) \qquad U_\kappa = \begin{pmatrix} F_\kappa & & & \\ & F_* & & \\ & & \ddots & \\ & * & & F_\kappa \end{pmatrix}$$

if the notation is chosen suitably.[8] The representations $F_1, F_2, \cdots, F_k$ are all distinct, and there are no other irreducible representations of $\mathfrak{G}$ in $\bar{K}$. Further, the F_κ are absolutely irreducible.

We denote the degree of F_κ by f_κ, that of U_κ by u_κ. Then U_κ appears f_κ times as indecomposable constituent of $\bar{R}$ and F_κ appears u_κ times as irreducible constituent of $\bar{R}$.

Let $c_{\kappa\lambda}$ be the multiplicity of F_λ as irreducible constituent of U_κ

$$(4) \qquad U_\kappa \leftrightarrow \sum_\lambda c_{\kappa\lambda} F_\lambda.$$

Here, the $c_{\kappa\lambda}$ are rational integers $\geqq 0$, the Cartan invariants[9] of $\mathfrak{G}$ (for p). They also can be characterized by means of structural properties of $\bar{\Gamma}$; they express mutual relations between the different prime ideal divisors of $\mathfrak{p}$ in $\mathfrak{J}$. Between the decomposition numbers and the Cartan invariants, we have the following equations[10]

$$(5) \qquad c_{\kappa\lambda} = \sum_{i=1}^{n} d_{i\kappa} d_{i\lambda} \qquad (\kappa, \lambda = 1, 2, \cdots, k)$$

or in matrix form

$$(6) \qquad C = D'D$$

where $C = (c_{\kappa\lambda})$, $D = (d_{i\kappa})$ and D' is the transpose of D.

There exists a representation (U_κ) of $\mathfrak{G}$ in K which if taken (mod $\mathfrak{p}$) becomes similar to U_κ, $\overline{(U_\kappa)} = U_\kappa$. We then have[11]

$$(7) \qquad (U_\kappa) \leftrightarrow \sum_j d_{j\kappa} Z_j.$$

6. Characters. Let M be a representation of $\mathfrak{G}$ which represents the group element G by $M(G)$. We denote the trace of the matrix $M(G)$ by $\chi(G)$. Then $\chi(G)$ is a function of the arbitrary group element G, the *character* of M. The

[8] See R. Brauer and C. Nesbitt, Nat. Ac. of Sciences 23, 1937, p. 236; C. Nesbitt, Ann. of Math. 39, 1938, p. 634. Free places in matrices are to be replaced by 0, the * stand for quantities in which we are not interested.

[9] E. Cartan, Annales de Toulouse 12B, 1898, p. 1.

[10] Three different proofs are given in M.R. pp. 9–11; T. Nakayama, Ann. of Math. 39, 1938, p. 361; R.A. pp. 257–258.

[11] Cf. R. A. The use of this fact which has not been mentioned in M.R. can be avoided, see footnote 13.

value of χ is the same for conjugate elements of $\mathfrak{G}$. We may, therefore, consider χ as a function of the classes of conjugate elements $\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_n$; we set $\chi_\nu = \chi(\mathfrak{C}_\nu) = \chi(G)$ if G belongs to $\mathfrak{C}_\nu$.

Let N be a second representation of $\mathfrak{G}$ with the character φ. We consider first the case that we have a ground field K of characteristic 0. If the determinants of $M(G)$ and $N(G)$ do not vanish, then the characters are equal, $\chi = \varphi$, if and only if M and N have the same irreducible constituents. Because of the full reducibility, this is the same as similarity of M and N. If we admit matrices of determinant 0 in M and N, we must add the assumption that M and N have the same degree. Otherwise, the (0)—representation may appear with different multiplicities in M and N.[12]

In the case of a ground field K of characteristic p, these theorems are not true. However, the method by which they are proved allows one to show that M and N have the same irreducible constituents, if and only if $M(G)$ and $N(G)$ have the same characteristic roots for every G in $\mathfrak{G}$.

We may write G as a product AB of two commutative elements where A has an order prime to p, whereas B has an order p^β, $\beta \geq 0$. The characteristic roots of $M(B)$ are all 1, being p^β-th roots of unity in a field of characteristic p. It follows that $M(G)$ and $M(A)$ have the same characteristic roots. It is, therefore, sufficient to require above that $M(G)$ and $N(G)$ have the same characteristic roots for every G of an order prime to p. Then the same will be automatically true for all G in $\mathfrak{G}$. We call an element G of $\mathfrak{G}$ *p-regular* if its order is prime to p.

We use the same notations as in §2. We set

$$g = p^a \cdot g' \qquad (p, g') = 1. \tag{8}$$

Let K_1 be the field obtained from K by the adjunction of the g'-th roots of unity $1, \delta, \delta^2, \cdots, \delta^{g'-1}$, let $\mathfrak{p}_1$ be a prime ideal divisor of $\mathfrak{p}$ in K_1, and let $\bar{K}_1$ be the field of integers of K_1 taken mod $\mathfrak{p}_1$. Then $\bar{K}_1$ is an extension field of $\bar{K}$, which contains the modular g'-th roots of unity $1, \bar{\delta}, \bar{\delta}^2, \cdots, \bar{\delta}^{g'-1}$, the residue classes of $1, \delta, \cdots \delta, \delta^{g'-1}$ (mod $\mathfrak{p}_1$). We have a $(1 - 1)$ relation between the ordinary and the modular g'-th roots of unity since $\delta^\alpha \not\equiv \delta^\beta$ (mod $\mathfrak{p}_1$) if $\delta^\alpha \neq \delta^\beta$.

If F now is a modular representation of $\mathfrak{G}$ with coefficients in $\bar{K}$ or in an extension field of $\bar{K}$, the characteristic roots of $F(G)$ will lie in $\bar{K}_1$. Let G be a p-regular element of $\mathfrak{G}$. We replace each such root $\bar{\delta}^\nu$ by δ^ν, and define now $\chi(G)$ as the sum of these δ^ν. In this manner, the character $\chi(G)$ is defined as a complex number for the p-regular elements G; the original value was the residue class $\bar{\chi}(G)$ of $\chi(G)$ (mod $\mathfrak{p}$). It now follows easily that two modular representations (with coefficients in $\bar{K}$ or in an extension field of $\bar{K}$) have the same irreducible constituents if and only if the two characters in the new sense coincide for p-regular elements.

[12] G. Frobenius and I. Schur, Sitzungsber. Preuss. Akad. 1906, p. 1906, p. 209.

7. The character relations. Let $\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_{k'}$ be the classes of conjugate elements which contain the p-regular elements. We denote by $\eta^{(\kappa)}$ the character of U_κ, by $\varphi^{(\kappa)}$ that of F_κ (cf. §5), by $\zeta^{(i)}$ that of Z_i. The value of a character for the class $\mathfrak{C}_\nu$ will be indicated by a suffix ν, e.g. $\eta_\nu^{(\kappa)}$, $(\nu = 1, 2, \cdots, k')$. The relations (7) and (2) now give[13]

$$\eta^{(\kappa)} = \sum_i d_{i\kappa} \zeta^{(i)} \tag{9}$$

$$\zeta^{(i)} = \sum_\lambda d_{i\lambda} \varphi^{(\lambda)} \tag{10}$$

$(i = 1, 2, \cdots, n;\ \kappa = 1, 2, \cdots, k)$. From these and (5), or directly from (4) we have

$$\eta^{(\kappa)} = \sum_\lambda c_{\kappa\lambda} \varphi^{(\lambda)}. \tag{11}$$

In particular, for the degrees u_κ, z_i, f_λ of U_κ, Z_i, F_λ, respectively, (9), (10), (11) for the unit element give

$$u_\kappa = \sum_i d_{i\kappa} z_i, \qquad z_i = \sum_\lambda d_{i\lambda} f_\lambda, \qquad u_\kappa = \sum_\lambda c_{\kappa\lambda} f_\lambda \tag{12}$$

since $u_\kappa = \eta^{(\kappa)}(1)$, $z_i = \zeta^{(i)}(1)$, $f_\lambda = \varphi^{(\lambda)}(1)$. We arrange $\varphi_\lambda^{(\kappa)}$, $\eta_\lambda^{(\kappa)}$ $\zeta_\lambda^{(i)}$ in matrix form

$$\Phi = (\varphi_\lambda^{(\kappa)}), \qquad \mathrm{H} = (\eta_\lambda^{(\kappa)}), \qquad \mathrm{Z} = (\zeta_\lambda^{(i)})$$

(κ row index, λ column index in Φ, H; i row index, λ column index in Z; $\kappa = 1, 2, \cdots, k$; $\lambda = 1, 2, \cdots, k'$; $i = 1, 2, \cdots, n$). Then relations (9), (10) and (11) become

$$\mathrm{H} = D'\mathrm{Z}, \qquad \mathrm{Z} = D\Phi, \qquad \mathrm{H} = C\Phi. \tag{13}$$

From the orthogonality relations for the ordinary group characters, we obtain

$$\mathrm{Z}'\mathrm{Z} = (g/g_\kappa \delta_{\kappa\lambda^*}) = \mathrm{T} \tag{14}$$

where g_κ denotes the number of elements in the class $\mathfrak{C}_\kappa$, and where the class $\mathfrak{C}_{\kappa^*}$ contains the elements reciprocal to those of $\mathfrak{C}_\kappa$ so that $1^*, 2^*, \cdots, k'^*$ is a permutation of $1, 2, \cdots, k'$. Then (14), (13) and (6) yield

$$\mathrm{H}'\Phi = \Phi' C \Phi = \mathrm{T}. \tag{15}$$

The equation (15) contains in matrix form orthogonality relations for the modular group characters, viz.

$$\sum_\rho \eta_\nu^{(\rho)} \varphi_\mu^{(\rho)} = \sum_{\rho,\sigma} \varphi_\nu^{(\rho)} c_{\rho\sigma} \varphi_\mu^{(\sigma)} = g\delta_{\nu\mu^*}/g_\nu \tag{16}$$

(see also relations (20), (21) and (22) below).

[13] We may avoid the use of (7) here by first deriving (10) from (2) and then (9) from (4), (5) and (10), see M.R.

8. Corollaries. Since T in (15) is non-singular, the columns of the matrix Φ which is of type (k, k') [14] are linearly independent, and hence $k \geqq k'$. On the other hand, the rows are linearly independent (mod $\mathfrak{p}$) because a linear relation would give a linear relation among the characters of $F_1, F_2, \cdots, F_k$, the values of these characters being understood as numbers of $\bar{K}$, as at the beginning of §6. Such a relation is impossible, hence $k = k'$. The number of distinct absolutely irreducible modular representations is equal to the number of classes of conjugate p-regular elements in $\mathfrak{G}$.[15] Further the determinant $|\Phi|$ of Φ is prime to $\mathfrak{p}$. Since $|\Phi|$ is integral, and, its square is rational according to (15) we see that $|\Phi|$ is prime to p.

$$(|\Phi|, p) = 1. \tag{17}$$

The column of H corresponding to the unit element of $\mathfrak{G}$ consists of $u_1, u_2, \cdots, u_k$. Since here $g_\nu = 1$, we obtain from (16) and (17) Dickson's theorem:[16]

$$u_\kappa \equiv 0 \quad (\text{mod } p^a), \qquad (\kappa = 1, 2, \cdots, k). \tag{18}$$

We have also now that all matrices which appear in (15) have inverses. Let us set, in particular, $C^{-1} = (\gamma_{\kappa\lambda})$. It follows from (15) that

$$\Phi \mathrm{T}^{-1} \mathrm{H}' = (\delta_{\kappa\lambda}) \tag{19}$$

which gives the character relations

$$\sum_\nu g_\nu \varphi_\nu^{(\kappa)} \eta_{\nu^*}^{(\lambda)} = \delta_{\kappa\lambda} g \qquad (\kappa, \lambda = 1, 2, \cdots, k). \tag{20}$$

In addition, multiplying (19) through by $C^{-1} = (\gamma_{\kappa\lambda})$ and using (13), we have

$$\Phi \mathrm{T}^{-1} \Phi' = C^{-1}, \text{ that is, } \sum_\nu g_\nu \varphi_\nu^{(\kappa)} \varphi_{\nu^*}^{(\lambda)} = \gamma_{\kappa\lambda} g \qquad (\kappa, \lambda = 1, 2, \cdots, k) \tag{21}$$

and from (19) multiplied through by C

$$\mathrm{H}\mathrm{T}^{-1}\mathrm{H}' = C, \text{ that is, } \sum_\nu g_\nu \eta_\nu^{(\kappa)} \eta_{\nu^*}^{(\lambda)} = c_{\kappa\lambda} g \qquad (\kappa, \lambda = 1, 2, \cdots, k). \tag{22}$$

If $(p, g) = 1$, then we have full reducibility, $U_\kappa = F_\kappa$. The matrix C, and then also D, is equal to the unit matrix, and we have $\bar{Z}_i = F_i$ (Speiser[17]).

9. Blocks. It is well known and easy to prove that the n elements

$$\Omega_\nu = \sum_{G \text{ in } \mathfrak{C}_\nu} G \qquad (\nu = 1, 2, \cdots, n) \tag{23}$$

form a basis of the centre of the group ring. Each irreducible representation of $\mathfrak{G}$ represents Ω_ν by a scalar multiple of the unit matrix I. We see

$$Z_i(\Omega_\nu) = \omega_\nu^{(i)} I, \qquad F_\kappa(\Omega_\nu) = \psi_\nu^{(\kappa)} I \tag{24}$$

[14] By a matrix of type (a, b), we understand a matrix with a rows and b columns.

[15] See R. Brauer, Actual. Scient. 195, Paris, 1935; M.R.

[16] See footnote 6, also M.R.

[17] Cf. A. Speiser, Theorie der Gruppen von endlicher Ordnung, 3rd edition, Berlin 1937, p. 223.

where $\omega_\nu^{(i)}$ is an integer of K, and $\psi_\nu^{(\kappa)}$ lies in $\bar{K}$. We say that F_κ and F_λ belong to the same *block*, if $\psi_\nu^{(\kappa)} = \psi_\nu^{(\lambda)}$ for $\nu = 1, 2, \cdots, n$. Then F_κ and F_λ represent the centre of $\bar{\Gamma}$ essentially in the same manner. Thus $F_1, F_2, \cdots, F_k$ appear distributed into s "*blocks*" $B_1, B_2, \cdots, B_s$.

We also speak of the U_κ which belong to a block B_τ by counting U_κ in B_τ if F_κ belongs to B_τ. Each matrix $U_\kappa(\Omega_\nu)$ can have only one characteristic root[18] which necessarily is $\psi_\nu^{(\kappa)}$ since F_κ is a constituent of U_κ, cf. (3). It follows that all the irreducible constituents of U_κ belong to B_τ. More generally, if in the sequence

$$U_\kappa, U_\alpha, U_\beta, \cdots, U_\sigma, U_\lambda \tag{25}$$

any two consecutive U_ρ have an irreducible constituent in common, then all the U_ρ and their irreducible constituents belong to the same block B_τ. If however U_κ and U_λ cannot be joined by such a chain (25), then it is easy to construct a centre element of $\bar{\Gamma}$ which is represented by I in F_κ and by 0 in F_λ so that F_κ and F_λ do not belong to the same block. We have here a new characterization of the blocks.

Assume now that $\bar{Z}_i$ contains F_κ as a irreducible constituent. From (24) it follows that

$$\bar{\omega}_\nu^{(i)} = \psi_\nu^{(\kappa)}$$

where the bar again indicates the residue class (mod $\mathfrak{p}$). All the irreducible constituents of $\bar{Z}_i$ belong necessarily to the same block B_τ. We now say that Z_i also belongs to the block B_τ. Two ordinary representations Z_i and Z_j belong to the same block if and only if $\omega_\nu^{(i)} \equiv \omega_\nu^{(j)}$ (mod $\mathfrak{p}$) for $\nu = 1, 2, \cdots, n$. Comparing the trace in the first formula (24) in a well known manner, we obtain

$$\omega_\nu^{(i)} = g_\nu \zeta_\nu^{(i)}/z_i \tag{26}$$

where z_i is the degree of Z_i. Hence, Z_i and Z_j belong to the same block if and only if

$$g_\nu \zeta_\nu^{(i)}/z_i \equiv g_\nu \zeta^{(j)}/z_j \pmod{\mathfrak{p}} \qquad (\nu = 1, 2, \cdots, n). \tag{27}$$

In what follows we shall always take $\varphi^{(1)}$, $\zeta^{(1)}$ to be the character of the unit representation considered as a modular and as an ordinary irreducible representation, respectively, of $\mathfrak{G}$, and B_1 to be the block which contains these characters.

We arrange the $F_1, F_2, \cdots, F_k$ and $Z_1, Z_2, \cdots, Z_n$ such that we first take the representations of B_1, then those of B_2, etc. Let x_τ be the number of Z_i belonging to B_τ and y_τ the number of F_i belonging to B_τ. It follows that C and D break up

$$C = \begin{pmatrix} C_1 & 0 & \cdots & 0 \\ 0 & C_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & C_s \end{pmatrix} \qquad D = \begin{pmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & D_s \end{pmatrix} \tag{28}$$

[18] Cf. R. Brauer and I. Schur, Sitzungsber. Preuss. Akad. 1930, p. 209, §2.

where C_τ is a square matrix of degree y_τ, and D_τ is of type (x_τ, y_τ). It is impossible to arrange the representations in such a manner that C_τ or D_τ break up further. For C_τ this follows directly from the properties of blocks given above. But because of (6), we have

$$C_\tau = D'_\tau D_\tau. \tag{29}$$

A breaking up of D_τ would imply one of C_τ. Since C_τ is non-singular (cf. (15) and (14)), we must have

$$x_\tau \geqq y_\tau. \tag{30}$$

We now form the trace of the element $F_\kappa(\Omega_\nu)$ in (24) and find $g_\nu\varphi_\nu^{(\kappa)}$, since Ω_ν is the sum of g_ν elements all of which have the trace $\varphi_\nu^{(\kappa)}$ in the representation F_κ. On the other hand, the trace of $F_\kappa(\Omega_\nu)$ is $f_\kappa\psi_\nu^{(\kappa)}$. If F_κ appears as modular constituent of the ordinary representation Z_i, then $\psi_\nu^{(\kappa)}$, as we have seen is the residue class of $\omega_\nu^{(i)}$ (mod $\mathfrak{p}$). Moreover, $\omega_\nu^{(i)}$ (mod $\mathfrak{p}$) depends only on the block B_τ to which F_κ and Z_i belong. We indicate that by setting $\omega_\nu^{(i)} \equiv \theta_\nu^{(\tau)}$ (mod $\mathfrak{p}$), where $\theta^{(\tau)}$ depends only on τ and not on i. Hence

$$g_\nu\varphi_\nu^{(\kappa)} \equiv f_\kappa\theta_\nu^{(\tau)} \pmod{\mathfrak{p}}. \tag{31}$$

Let us set $gC^{-1} = (\tilde{\gamma}_{\kappa\lambda})$, that is, $\tilde{\gamma}_{\kappa\lambda} = g\gamma_{\kappa\lambda}$. From (21) it follows that the $\tilde{\gamma}_{\kappa\lambda}$ are algebraic integers; that they are rational comes as a consequence of their definition. From (31) and (21) we have

$$\tilde{\gamma}_{\kappa\lambda} \equiv f_\kappa \sum_{\nu=1}^{k} \theta_\nu^{(\tau)} \varphi_{\nu^*}^{(\lambda)} \pmod{\mathfrak{p}} \tag{32}$$

The sum on the right depends on τ and λ; we denote it by $S(\tau, \lambda)$. If F_λ also belongs to the block B_τ, then by reasons of symmetry

$$\tilde{\gamma}_{\kappa\lambda} \equiv f_\kappa S(\tau, \lambda) \equiv f_\lambda S(\tau, \kappa) \pmod{\mathfrak{p}}. \tag{33}$$

In particular, if f_κ or f_λ are divisible by p, then $\tilde{\gamma}_{\kappa\lambda}$ is divisible by $\mathfrak{p}$ and hence $\tilde{\gamma}_{\kappa\lambda} \equiv 0 \pmod{p}$. If $f_\kappa \not\equiv 0 \pmod{p}$, then (33) shows that the value of $S(\tau, \kappa)/f_\kappa$ (mod $\mathfrak{p}$) depends only on τ. We may, therefore, write

$$\tilde{\gamma}_{\kappa\lambda} \equiv f_\kappa f_\lambda S_\tau \pmod{\mathfrak{p}}$$

where S_τ depends only on τ. We may here take S_τ as a rational integer and have

$$\tilde{\gamma}_{\kappa\lambda} \equiv f_\kappa f_\lambda S_\tau \pmod{p} \qquad \text{if } F_\kappa \text{ and } F_\lambda \text{ in } B_\tau. \tag{34}$$

If F_κ and F_λ belong to different blocks, then $\tilde{\gamma}_{\kappa\lambda} = 0$, because of the form (28) of C.

Since $\varphi^{(1)}$ is the character of the 1-representation and is contained in the block B_1, then (21) and (34) show

$$N = \tilde{\gamma}_{11} \equiv S_1 \pmod{p} \tag{35}$$

where N is the number of all p-regular elements of $\mathfrak{G}$.

10. Decomposition of Γ. The block properties derived in §9 are those which we are going to use in the later sections. But the importance of these blocks can better be recognized from other facts which we shall describe briefly.[19] Let B_τ be a fixed block, and consider the elements $\bar{\alpha}$ of $\bar{\Gamma}$, for which $U_\kappa(\bar{\alpha}) = 0$ for every U_κ except for those of B_τ. These $\bar{\alpha}$ form an invariant subalgebra $\sum_\tau$, and we have

$$\bar{\Gamma} = \sum\nolimits_1 \oplus \sum\nolimits_2 \oplus \cdots \oplus \sum\nolimits_s .$$

The $\sum_\tau$ cannot be represented as direct sums.

In close connection with this fact, we have the following ideal theoretical significance of the blocks. The ideal $\mathfrak{p}$ of $\mathfrak{J}$ (cf. §2) can be written uniquely as the intersection of s ideals $\mathfrak{M}_\tau$ ($\neq \mathfrak{J}$) any two of which are relatively prime

$$\mathfrak{p} = [\mathfrak{M}_1, \mathfrak{M}_2, \cdots, \mathfrak{M}_s]$$

and $\mathfrak{M}_\tau$ cannot be written as intersection of relatively prime ideals $\neq \mathfrak{M}_\tau$. There are exactly k prime ideal divisors $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_k$ of $\mathfrak{p}$ in $\mathfrak{J}$. Here $\mathfrak{P}_\kappa$ can be defined as the set of those elements α of $\mathfrak{J}$ for which $F_\kappa(\alpha) = 0$. Two representations F_κ and F_λ belong to the same block if and only if $\mathfrak{P}_\kappa$ and $\mathfrak{P}_\lambda$ divide the same $\mathfrak{M}_\tau$.

11. Summary of the results. The principal aim of this paper is the proof of the following two theorems.

THEOREM 1. *Let $\mathfrak{G}$ be a group of order $g = p^a g'$, p a prime, $(g', p) = 1$. An ordinary irreducible representation Z_i of a degree $z_i \equiv 0 \pmod{p^a}$ remains irreducible as a modular representation, i.e. $\bar{Z}_i$ is equal to one of the F_κ, and $U_\kappa = F_\kappa$. Further Z_i forms a block B_τ of its own. The character $\zeta^{(i)}$ of Z_i vanishes for all elements of an order divisible by p.*

We denote a block B_τ of this kind as a block of highest kind. In the notations used above, we have here $x_\tau = y_\tau = 1$. We also shall show that for the blocks which are not of highest kind, we have $x_\tau > y_\tau$, in particular, $x_\tau > 1$.

THEOREM 2. *Let t_0 be the number of classes $\mathfrak{C}_\nu$ of conjugate elements in $\mathfrak{G}$ such that* (a) *the number of elements in $\mathfrak{C}_\nu$ is prime to p,* (b) *the elements of $\mathfrak{C}_\nu$ have an order prime to p. There exist exactly t_0 blocks B_τ which contain at least one ordinary irreducible representation Z_i of a degree z_i prime to p.*

We denote blocks of the type mentioned in this theorem as blocks of lowest kind. We also obtain some results for the blocks of intermediate types α, which contain only Z_i of degree $z_i \equiv 0 \pmod{p^\alpha}$ such that at least one of these degrees $z_i \not\equiv 0 \pmod{p^{\alpha+1}}$. The method which yields theorem 1 can, in a far more elaborate form, be used for a study of the blocks of type $a - 1$ as will be

---

[19] See R. A.

shown in another paper.[20] In the case $a = 1$, i.e. $g = p \cdot g'$, $(p, g') = 1$, each block is either of the highest or of lowest kind, so that the results give some information about every block. This can be made the basis for a study of this class of groups, which yields a large number of new results.[20] In order to attack the general group of finite order in a similar manner it would be necessary first to refine greatly the theory of blocks.

Two of the most important tools for the computation of the ordinary group characters are formed by the method of the multiplication of characters and the Frobenius' method of constructing characters of $\mathfrak{G}$ from characters of a subgroup $\mathfrak{H}$ of $\mathfrak{G}$. These methods can also be applied to modular characters. In part IV we study the former method, and in part V, the Frobenius' method. In part VI we consider a number of special cases and examples. We hope that in the results of these latter parts of our paper some justification can be seen for the somewhat complicated theory as developed in this lengthy introduction.

II. Blocks of Highest Kind

12. Condition for the reducibility of $\bar{Z}_i$. We use the same notations as in the introduction. If for one of the Z_i (§4) the corresponding modular representation $\bar{Z}_i$ becomes reducible, then there exists a non-singular matrix $\bar{M} = (\bar{m}_{ij})$ in the field $\bar{K}$ such that $\bar{M}^{-1}\bar{Z}_i\bar{M}$ breaks up into at least two constituents

$$\bar{M}^{-1}\bar{Z}_i\bar{M} = \begin{pmatrix} \bar{W}_1 & 0 \\ \bar{W}^3 & \bar{W}_4 \end{pmatrix}.$$

We choose a matrix $M = (m_{ij})$ such that m_{ij} lies in the residue class $\bar{m}_{ij}$ (mod $\mathfrak{p}$). Then the determinant of M is prime to $\mathfrak{p}$ and hence different from 0. Forming $M^{-1}Z_iM$, we obtain a formula

$$Z_i^* = M^{-1}Z_iM = \begin{pmatrix} W_1 & W_2 \\ W_3 & W_4 \end{pmatrix} \tag{36}$$

where all the coefficients in W_1, W_2, W_3, W_4, are $\mathfrak{p}$-integers of K, and those of W_2 are divisible by $\mathfrak{p}$ (i.e. each coefficient in W_2 is the quotient α/β of two integers of K such that $\alpha \equiv 0$, $\beta \not\equiv 0$ (mod $\mathfrak{p}$)). Let $Z_i^*(G) = (w_{\kappa\lambda}^{(i)}(G))$. According to the formulas of I. Schur,[21] we have[22]

$$\sum_G w_{\kappa\lambda}^{(i)}(G)w_{\rho\sigma}^{(i)}(G^{-1}) = g/z_i\,\delta_{\kappa\sigma}\delta_{\lambda\rho} \tag{37}$$

$$\sum_G w_{\kappa\lambda}^{(i)}(G)w_{\rho\sigma}^{(j)}(G^{-1}) = 0 \qquad \text{for } i \neq j. \tag{38}$$

In (37) we now take $\kappa = \sigma = 1$, $\lambda = \rho = z_i$ so that $\delta_{\kappa\sigma} = \delta_{\lambda\rho} = 1$. From the form of W_2 in (36), we have $w_{1z_i}(G) \equiv 0$ (mod $\mathfrak{p}$) for every G, hence $g/z_i \equiv 0$ (mod $\mathfrak{p}$) and consequently $g/z_i \equiv 0$ (mod p). Since g was exactly divisible by

[20] R. Brauer, Nat. Ac. of Sciences 25, 1939, p. 290.

[21] I. Schur, Sitzungsber. Preuss. Akad. 1905, p. 406.

[22] We set $\delta_{ij} = 0$ for $i \neq j$, $\delta_{ii} = 1$.

p^a (cf. (8)) we have $z_i \not\equiv 0 \pmod{p^a}$ if $\bar{Z}_i$ is reducible. This shows that if $z_i \equiv 0 \pmod{p^a}$, then $\bar{Z}_i$ is irreducible. This was the first part of theorem 1.

13. Blocks of highest kind. We now have to show that a Z_i with $z_i \equiv 0 \pmod{p^a}$ forms a block of its own. If this were not so, then we would have a Z_j with $i \neq j$ such that $\bar{Z}_i$ and $\bar{Z}_j$ have an irreducible constituent in common. But since $\bar{Z}_i$ itself is irreducible, $\bar{Z}_i$ would have to occur as constituent of $\bar{Z}_j$. Then there would exist a matrix $\bar{L} = (l_{ij})$ with coefficients in $\bar{K}$ such that[23]

$$\bar{L}^{-1}\bar{Z}_j\bar{L} = \begin{pmatrix} \bar{W}_1 & 0 & 0 \\ \bar{W}_4 & \bar{Z}_i & 0 \\ \bar{W}_7 & \bar{W}_8 & \bar{W}_9 \end{pmatrix}.$$

Choosing again the element l_{ij} in the residue class $l_{ij} \pmod{\mathfrak{p}}$, and setting $L = (l_{ij})$, we then have a formula

$$L^{-1}Z_jL = \begin{pmatrix} W_1 & W_2 & W_3 \\ W_4 & W_5 & W_6 \\ W_7 & W_8 & W_9 \end{pmatrix} \tag{39}$$

where

$$W_5 \equiv Z_i \pmod{\mathfrak{p}}. \tag{40}$$

We choose now $\kappa = \lambda = 1$ in (38), for ρ we take the number of the first row of W_5 in (39), for σ the number of the first column of W_5. Then (38) yields because of (40) and (37).

$$0 = \sum w_{11}^{(i)}(G)w_{\rho\sigma}^{(j)}(G^{-1}) \equiv \sum w_{11}^{(i)}(G)w_{11}^{(i)}(G^{-1}) = g/z_i \pmod{\mathfrak{p}}.$$

But this is impossible, because $z_i \equiv 0 \pmod{p^a}$. Consequently, no $\bar{Z}_j$ with $j \neq i$ belongs to the same block B_τ as Z_i. The block B_τ contains only the one ordinary irreducible representation Z_i, and only the one modular irreducible representation $\bar{Z}_i = F_\kappa$. In (28), C_τ and D_τ are matrices of degree 1, $x_\tau = y_\tau = 1$, and we have $D_\tau = 1$. Hence $C_\tau = 1$, because of (29). From (4), we obtain $U_\kappa = F_\kappa$.

14. Vanishing of the character for p-singular elements of $\mathfrak{G}$. Let H be a fixed element of $\mathfrak{G}$ the order of which is divisible by p. From the orthogonality relations for ordinary group characters, it follows that we have

$$\sum_{i=1}^{n} \zeta^{(i)}(G)\zeta^{(i)}(H) = 0$$

for every p-regular element G, since G and H^{-1} cannot be conjugate in $\mathfrak{G}$. Using (10), this can be written in the form

$$\sum_{i=1}^{n} \sum_{\kappa=1}^{k} d_{i\kappa}\varphi^{(\kappa)}(G)\zeta^{(i)}(H) = 0. \tag{41}$$

[23] The first row and column on the right side may be missing, also the last row and column.

Since (41) represents a linear relation between $\varphi^{(1)}(G)$, $\varphi^{(2)}(G)$, $\cdots$, $\varphi^{(k)}(G)$ which is true for every non-singular element G, the coefficient of each $\varphi^{(\kappa)}(G)$ must vanish

$$\sum_{i=1}^{n} d_{i\kappa}\zeta^{(i)}(H) = 0 \qquad (\kappa = 1, 2, \cdots, k).$$

If $\bar{Z}_i = F_\kappa$ as in §13, then $d_{i\kappa} = 1$, and $d_{j\kappa} = 0$ for $i \neq j$ since F_κ does not appear in $\bar{Z}_j$. Hence $\zeta^{(i)}(H) = 0$. This proves theorem 1 completely.

In the case $(g, p) = 1$, it follows at once from theorem 1 that each ordinary irreducible representation Z_i remains irreducible when taken as a modular representation and so $\bar{Z}_i$ is equivalent to some F_κ. Then $k = n$ and the $\zeta^{(i)}$ are identical with the $\varphi^{(\kappa)}$, $\eta^{(\kappa)}$. The relations (16) and (20) here become the same as the Frobenius relations for group characters (cf. §§3, 8.)

15. Example. As an example, we mention the simple Mathieu group M_{12} of order $12.11.10.9.8 = 2^6.3^3.5.11 = 95040$, the characters of which have been given by Frobenius.[24] The degree of the ordinary irreducible representations are

1, 11, 11, 16, 16, 45, 54, 55, 55, 55, 66, 99, 120, 144, 176.

Of these 15 characters, 8 are of highest kind (mod 11), for instance, the character of degree 176. From the table of characters, it follows that this character is the product of a character of degree 11 and a character of degree 16. Consequently the characters of degree 16 must also remain irreducible (mod 11), since a splitting would imply a splitting of the character of degree 176. The characters of degree 16 are not of highest kind.

For $p = 5$ we have 5 characters of highest kind, for $p = 3$ there is one of them and for $p = 2$ there is no such character.

III. The Elementary Divisors of C

16. Computation of the elementary divisors of C. In the following section we work in the ring $\mathfrak{o}_\mathfrak{p}$ of $\mathfrak{p}$-integers of K. If π is an element such that $\pi \equiv 0 \pmod{\mathfrak{p}}$, $\pi \not\equiv 0 \pmod{\mathfrak{p}^2}$, then every ideal of $\mathfrak{o}_\mathfrak{p}$ is of the form $(\pi)^m$, and therefore, the theory of elementary divisors holds for matrices with coefficients in $\mathfrak{o}_\mathfrak{p}$. In formula (15), $\Phi'C\Phi = \mathrm{T}$, the determinant of Φ is a unit of $\mathfrak{o}_\mathfrak{p}$ because of (17). Consequently, C and T have the same elementary divisors. But the elementary divisors of T can be obtained directly from (14), they are the highest powers of p which divide the numbers

$$(42) \qquad g/g_1, g/g_2, \cdots, g/g_k.$$

We now consider C as a matrix with coefficients in the ring of rational integers, and denote the elementary divisors corresponding to this case by $e_1, e_2, \cdots, e_k$. It follows that *the powers of p which divide these integers are exactly the same*

[24] The ordinary characters of the Mathieu-groups have been given by G. Frobenius, Sitzungsber. Preuss. Akad. 1904, p. 558.

powers which appear in the integers g/g_ν, $\nu = 1, 2, \cdots, k$, if the latter are properly arranged.

In another paper, it will be shown that the determinant of C is actually a power of p. Then it follows that the e_ν are themselves the powers of p which divide the numbers (42). For our present purpose this finer result will not be needed.

17. Blocks of type α. We say that a block B_τ is of type α, if it contains only representations F_κ of degrees $f_\kappa \equiv 0 \pmod{p^\alpha}$, and if at least one of these degrees is not divisible by $p^{\alpha+1}$.

By (12)

$$u_\kappa = \sum_\lambda c_{\kappa\lambda} f_\lambda \qquad (\lambda = 1, 2, \cdots, k).$$

Because of the form (28) of C, the corresponding formulas hold, when we restrict κ and λ to those values for which F_κ, F_λ belong to the block B_τ.

We can find two unimodular matrices M_1 and M_2 such that $C_\tau^* = M_1 C_\tau M_2$ has zeros outside of the main diagonal, and contains the elementary divisors e_κ in the main diagonal.[25] We have then

$$u_\kappa^* = e_\kappa f_\kappa^* \tag{43}$$

where the u_κ^* are obtained from the u_λ by the linear transformation M_1, and the f_κ^* from the f_λ by the transformation M_2^{-1}; κ, λ range over the values corresponding to B_τ. Since M_2 is unimodular, all the f_κ^* are divisible by p^α, and one of them is not divisible by $p^{\alpha+1}$. On the other hand, the u_κ^* are divisible by p^a, according to (18). From (43) it follows that *at least one of the elementary divisors e_κ corresponding to B_τ is divisible by $p^{a-\alpha}$.* Let s_α denote the number of blocks of type α, and a_α the number of integers g_ν ($\nu = 1, 2, \cdots, k$) which are divisible by p^α and not by $p^{\alpha+1}$. We consider now the $s_\alpha + s_{\alpha-1} + \cdots + s_0$ blocks of type $\leqq \alpha$. To each of them corresponds an elementary divisor which is divisible at least by $p^{a-\alpha}$ and hence a number g/g_ν which is at least divisible by this number. Then g_ν will be divisible at most by p^α, and we find

$$s_0 + s_1 + \cdots + s_\alpha \leqq a_0 + a_1 + \cdots + a_\alpha .$$

THEOREM 3. *Let $\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_k$ be the classes of conjugate p-regular elements in $\mathfrak{G}$ and denote by g_ν the number of elements in $\mathfrak{C}_\nu$. If a_α of the numbers g_ν are divisible by p^α and not by $p^{\alpha+1}$, and if $\mathfrak{G}$ possesses s_α blocks of type α, then* (for $\alpha = 0, 1, 2, \cdots, a$)

$$s_0 + s_1 \cdots + s_\alpha \leqq a_0 + a_1 \cdots + a_\alpha . \tag{44}$$

18. Ordinary characters which are linearly independent mod $\mathfrak{p}$.

LEMMA: There exist k but not more than k ordinary irreducible characters $\zeta^{(i)}$ which are linearly independent (mod $\mathfrak{p}$).

[25] The elementary divisors of all the C_τ together are $e_1, e_2, \cdots, e_k$ (in some arrangement) because of (28).

Proof: If a linear relation

$$\sum_i a_i \zeta^{(i)} \equiv 0 \pmod{\mathfrak{p}}$$

with p-integral coefficients holds for the p-regular classes, then it holds for every class. This can be seen similarly as in §6. From this remark, it follows already that we cannot have more than k ordinary characters which are linearly independent (mod $\mathfrak{p}$).

For the proof that there exist k such independent characters, we use a method by which one of us showed earlier the existence of k irreducible modular characters.[26] If the maximal number of ordinary irreducible characters, which are linearly independent mod $\mathfrak{p}$, was smaller than k, then we could find $\mathfrak{p}$-integers b_ν such that

$$\sum_{\nu=1}^{k} b_\nu \zeta_\nu^{(i)} \equiv 0 \pmod{\mathfrak{p}}$$

and the b_ν are not all divisible by $\mathfrak{p}$. Since a reducible character of $\mathfrak{G}$ is a sum of irreducible character $\zeta^{(i)}$, we would have

$$\sum_{\nu=1}^{k} b_\nu \zeta_\nu \equiv 0 \pmod{\mathfrak{p}} \tag{45}$$

for any character ζ of $\mathfrak{G}$. We want to show that this is impossible if the b_ν are not all divisible by $\mathfrak{p}$. We assume that the corresponding result for all proper subgroups $\mathfrak{H}$ of $\mathfrak{G}$ has already been shown.

Let $\mathfrak{H}$ be a proper subgroup of $\mathfrak{G}$, let $\mathfrak{C}_1', \mathfrak{C}_2', \cdots, \mathfrak{C}_l'$ be the classes of conjugate p-regular elements of $\mathfrak{H}$ and let H_γ be an element of $\mathfrak{C}_\gamma'$. If ψ is any character of $\mathfrak{H}$, we set $\psi(G) = 0$ if G does not belong to $\mathfrak{H}$. We determine a (right hand side) residue system $P_1, P_2, \cdots, P_m$ of $\mathfrak{G}$ mod $\mathfrak{H}$. According to Frobenius.[27]

$$\zeta(G) = \sum_{\mu=1}^{m} \psi(P_\mu G P_\mu^{-1}) \tag{46}$$

is a character of $\mathfrak{G}$. Since (45) hold for every character of $\mathfrak{G}$, it holds for (46). The elements G and $P_\mu G P_\mu^{-1}$ are p-regular at the same time. If G_ρ is an element of $\mathfrak{C}_\rho$, we have from (46)

$$\zeta_\rho = \sum_{\sigma=1}^{l} l_{\rho\sigma} \psi_\sigma \tag{47}$$

where $l_{\rho\sigma}$ denotes the number of P_μ for which $P_\mu G_\rho P_\mu^{-1}$ is conjugate to H_σ with regard to $\mathfrak{H}$. For each H_σ there exists exactly one ρ for which $l_{\rho\sigma} \neq 0$ since one class $\mathfrak{C}_\rho$ in $\mathfrak{G}$ must contain H_σ. We denote this ρ by $\tau(\sigma)$. From (45) and (47) it follows that

$$\sum_{\sigma=1}^{l} \left(\sum_{\rho=1}^{k} b_\rho l_{\rho\sigma} \right) \psi_\sigma \equiv 0 \pmod{\mathfrak{p}}. \tag{48}$$

[26] See footnote 15.

[27] G. Frobenius, Sitzungsber. Preuss. Akad. 1898, p. 501.

This must hold for every character ψ of $\mathfrak{H}$. But (48) represents a congruence of exactly the same type for $\mathfrak{H}$, as (45) has for $\mathfrak{G}$. According to our assumption concerning $\mathfrak{H}$, the coefficients of every ψ_σ must be divisible by $\mathfrak{p}$,

$$\sum_{\rho=1}^{k} b_\rho l_{\rho\sigma} \equiv 0 \pmod{\mathfrak{p}} \qquad \text{for } \sigma = 1, 2, \cdots, l$$

and since only $l_{\tau(\sigma),\sigma} \not\equiv 0$

$$b_{\tau(\sigma)} l_{\tau(\sigma),\sigma} \equiv 0 \pmod{\mathfrak{p}} \qquad \text{for } \sigma = 1, 2, \cdots, l. \tag{49}$$

So far the subgroup $\mathfrak{H}$ has been arbitrary. We now try to determine $\mathfrak{H}$ for a given value ρ, ($\rho = 1, 2, \cdots, k$), such that (a) ρ appears in the form $\rho = \tau(\sigma)$, and (b) for this σ the number $l_{\rho\sigma}$ is not divisible by $\mathfrak{p}$. Then (49) implies $b_\rho \equiv 0 \pmod{\mathfrak{p}}$, and if this holds for every ρ, then we have arrived at a contradiction with the fact that the congruence (45) was to be not trivial.

The condition (a) is satisfied when G_ρ belongs to $\mathfrak{H}$, we may take $G_\rho = H_\sigma$. If we choose $\mathfrak{H}$ as a subgroup of the normalizer $\mathfrak{N}$ of G_ρ, then G_ρ is only conjugate to itself with regard to $\mathfrak{H}$. Then $l_{\rho\sigma}$ in (46) can be defined as the number of P_μ for which $P_\mu G_\rho P_\mu^{-1} = G_\rho$. If $\mathfrak{H}$ has the order h, then $hl_{\rho\sigma} = N$ is the order of $\mathfrak{N}$. We have only to take care that $\mathfrak{H}$ contains a p-Sylow group of $\mathfrak{N}$. Then h is divisible by the same power of p as N, hence $l_{\rho\sigma} \not\equiv 0 \pmod{\mathfrak{p}}$ and, therefore, condition (b) is satisfied.

We can, therefore, satisfy the above conditions (a) and (b) by choosing $\mathfrak{H}$ as the subgroup which is generated by G_ρ and a p-Sylow group of the normalizer $\mathfrak{N}$ of G_ρ. Here, however, an exceptional case is possible which must be treated separately. The group defined in this manner can be identical with $\mathfrak{G}$.

In this case, the only p-regular elements of $\mathfrak{G}$ are $1, G_\rho, G_\rho, \cdots, G_\rho^{q-1}$, where q is the order of G_ρ. We obtain a character of $\mathfrak{G}$ by associating ϵ^μ with G^μ where ϵ is a q-th root of unity. Then (45) becomes

$$\sum_{\mu}^{q-1} b_\mu \epsilon^\mu \equiv 0 \pmod{\mathfrak{p}}.$$

We multiply here with $\epsilon^{-\beta}$ for a fixed β and add over all q-th roots of unity. Since $(q, p) = 1$ we find $b_\beta \equiv 0 \pmod{\mathfrak{p}}$ for $\beta = 0, 1, 2, \cdots, q-1$, which gives a contradiction.

19. Applications of the lemma. It follows immediately from the lemma in §18 that the congruences

$$\sum_{i=1}^{n} a_i \zeta_\nu^{(i)} \equiv \eta_\nu \pmod{\mathfrak{p}} \qquad (\text{for } \nu = 1, 2, \cdots, k) \tag{50}$$

can be solved with regard to $a_1, a_2, \cdots, a_n$ if $\eta_1, \eta_2, \cdots, \eta_k$ are any given $\mathfrak{p}$-integers of K. The a_i also will be $\mathfrak{p}$-integers of K.

From (10) and (28) the number of $\zeta^{(i)}$ in a given block B_τ which are linearly independent mod $\mathfrak{p}$, is at most equal to the number y_τ of modular characters

$\varphi^{(\kappa)}$ in B_τ. But since $y_1 + \cdots + y_s$ is the full number k of modular irreducible characters, this implies that B_τ contains y_τ characters $\zeta^{(i)}$ which are linearly independent (mod $\mathfrak{p}$). It follows that the matrix D_τ of type (x_τ, y_τ) in (28) still has the rank y_τ when it is considered mod $\mathfrak{p}$.

From this remark and (10) it follows that the modular characters $\varphi^{(\kappa)}$ can be expressed by means of the ordinary characters with p-integral rational coefficients. For a block of type α, all the f_κ are divisible by p^α. Since by (12), $z_i = \sum_\kappa d_{i\kappa} f_\kappa$, it follows that all the z_i of the block B_τ will be divisible by p^α. On the other hand, the z_i of B_τ cannot all be divisible by $p^{\alpha+1}$, since otherwise all the f_κ of B_τ would be divisible by $p^{\alpha+1}$, as we see when we express the $\varphi^{(\kappa)}$ of B_τ as linear combinations of the $\zeta^{(i)}(G)$ of B_τ with p-integral coefficients and set $G = 1$. *We can define a block B_τ of type α by the fact that the degrees of the ordinary irreducible characters of B_τ are all divisible by p^α but not by $p^{\alpha+1}$.* In the definition in §17 we can replace the modular characters by ordinary characters. In particular, the blocks of type 0 are the blocks of lowest kind; the blocks of type a, the blocks of highest kind (§11).

If the $\zeta^{(i)}$ of B_τ are arranged in a suitable order, then the first y_τ of them will be linearly independent mod $\mathfrak{p}$. We may then find a matrix V of degree y_τ with p-integral rational coefficients and a determinant prime to p such that

$$D_\tau V = \begin{pmatrix} I \\ M \end{pmatrix}$$

where M is a matrix of type $(x_\tau - y_\tau, y_\tau)$ and I, the unit matrix of degree y_τ. Using (29), we find

$$V'C_\tau V = V'D_\tau' D_\tau V = (I, M')\begin{pmatrix} I \\ M \end{pmatrix} = I + M'M. \tag{51}$$

We work in the Galois field with p elements, replacing every number by its residue class (mod p). If M in this sense has rank m, then we can find $y_\tau - m$ linearly independent vectors ξ of y_τ dimensions for which $M\xi = 0$. For these vectors, we have $(I + M'M)\xi = \xi$ so that at least $y_\tau - m$ linearly independent vectors are obtained in the form $(I + M'M)\eta$ where η is an arbitrary vector. It follows that $(I + M'M)$ has (mod p) a rank $r \geqq y_\tau - m$. Because of (51) C_τ has (mod p) the same rank r. Then exactly r of the elementary divisors of C_τ will be not divisible by p. But $m \leqq x_\tau - y_\tau$, since M has $x_\tau - y_\tau$ rows, so

$$r \geqq y_\tau - (x_\tau - y_\tau) = 2y_\tau - x_\tau.$$

THEOREM 4. *If the block B_τ contains x_τ ordinary and y_τ modular irreducible characters, then the corresponding part C_τ of the Cartan matrix has at least $2y_\tau - x_\tau$ elementary divisors which are not divisible by p (and hence equal to* 1 *according to the theorem quoted in* §16).

If $2y_\tau < x_\tau$ then this theorem does not give anything.

If for a block we have $y_\tau = x_\tau$, then x_τ of the elementary divisors will be prime to p. But this is the total number of elementary divisors of B_τ. Ac-

cording to §17, this is impossible if B_τ is of type $\alpha < a$. Hence B_τ is of highest type, and then $x_\tau = y_\tau = 1$ (theorem 1). Using (30) we obtain

THEOREM 5: *Every block, which is not of highest kind, contains more ordinary than modular irreducible characters.*[28]

This shows that Theorem 1 characterizes the blocks of highest kind.

20. Blocks of lowest kind. We now come to the proof of theorem 2 (§11). Let $\eta_1, \eta_2, \cdots, \eta_k$ be any given $\mathfrak{p}$-integers of K. We solve the congruences (50). Let $\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_m$ be these classes of p-regular elements of $\mathfrak{G}$, for which the number g_ν of elements in the class is not divisible by p, then

$$m = a_0 \tag{52}$$

where a_0 was defined in §17.

The number $\omega_\nu^{(i)} = g_\nu \zeta_\nu^{(i)}/z_i$ (cf. (24) and (26)) is an algebraic integer. If $z_i \equiv 0 \pmod p$ and $(g_\nu, p) = 1$, then $\zeta_\nu^{(i)} \equiv 0 \pmod{\mathfrak{p}}$. The corresponding terms in (50) can be omitted. We have, therefore,

$$\sum_i a_i \zeta_\nu^{(i)} \equiv \eta_\nu \pmod{\mathfrak{p}} \qquad \text{for } \nu = 1, 2, \cdots, m \tag{53}$$

where the sum is extended over such values of i, for which $z_i \not\equiv 0 \pmod p$. In particular, only characters of blocks of the lowest kind appear. We can pick out one character $\zeta^{(h)}$ in B_τ such that $z_h \not\equiv 0 \pmod p$. If $\zeta^{(i)}$ is another character of B_τ, then according to (26) and (27) we have

$$\frac{g_\nu \zeta_\nu^{(i)}}{z_i} \equiv \frac{g_\nu \zeta_\nu^{(h)}}{z_h} \pmod{\mathfrak{p}}$$

$$\zeta_\nu^{(i)} \equiv \frac{z_i}{z_h} \zeta_\nu^{(h)}, \pmod{\mathfrak{p}}, \qquad (\nu = 1, 2, \cdots, m)$$

since $(g_\nu, p) = 1$, for $\nu = 1, 2, \cdots, m$. We substitute this value in (53) and obtain formulae

$$\sum_h b_h \zeta_\nu^{(h)} \equiv \eta_\nu \pmod{\mathfrak{p}}, \qquad (\nu = 1, 2, \cdots, m) \tag{54}$$

where the $\zeta^{(h)}$ are the characters we selected in the blocks of lowest kind. The number of terms on the left side then is the number s_0 of blocks of the lowest kind. The b_h are $\mathfrak{p}$-integers which are independent of ν. For every given set of $\mathfrak{p}$-integers $\eta_1, \eta_2, \cdots, \eta_m$ the congruences have a solution b_h. The number of unknowns b_h cannot be smaller than the number of congruences, hence, from (52)

$$a_0 = m \leqq s_0 .$$

But from (44), it follows that $s_0 \leqq a_0$. Consequently, $s_0 = a_0$, and this is exactly the statement of theorem 2, §11.

[28] A second proof for this relation $x_\tau > y_\tau$ is given in §27. It also can be proved by considering the representation of the elements of the center of the group ring.

In other words this result can be expressed as follows: An elementary divisor divisible by p^a can appear in C_τ only if the block B_τ is of lowest kind. In this case there is exactly one such elementary divisor.

We easily see now that for blocks of type 0, 1, and a there is at least one U_κ of the block whose degree $u_\kappa \not\equiv 0 \pmod{p^{a+1}}$. For a block B_τ of type α it follows from (43) that if all $u_\kappa \equiv 0 \pmod{p^{a+1}}$ then at least one elementary divisor of C_τ is divisible by $p^{a-\alpha+1}$. For $\alpha = 0$ this would mean an elementary divisor divisible by p^{a+1} which is not possible (cf. §16). For $\alpha = 1$ it means an elementary divisor divisible by p^a but by the above statement of theorem 2 such divisors can appear only for blocks of type 0. In case $\alpha = a$ the remark is obvious, since the block is then of highest kind. For values of α intermediate to 1 and a we can only as yet say that an elementary divisor of C_τ is divisible at most by p^{a-1}, and hence from (43) at least one u_k is divisible at most by $p^{a+\beta}$ where $\beta \leqq \alpha - 1$.

21. Alternative proof of theorem 2. We here begin with the components $(\tilde{\gamma}_{\kappa\lambda})$ of the matrix gC^{-1} (cf. §9), and again work in the ring of all rational p-integers. If the matrix C_τ (cf. (28)) has the elementary divisors p^{α_ν}, $(\nu = 1, 2, \cdots, y_\tau)$, then gC_τ^{-1} has the elementary divisors $p^{a-\alpha_\nu}$. In the case that B_τ is not a block of the lowest kind, then for any pair F_κ, F_λ belonging to B_τ the degrees f_κ, f_λ are divisible by p, and so from (34) $\tilde{\gamma}_{\kappa\lambda} \equiv 0 \pmod p$. Hence all the α_ν are smaller than a in this case. If B_τ is of the lowest kind, then since by (34) the matrix $gC_\tau^{-1} \equiv (f_\kappa f_\lambda S_\tau)$, ($\kappa$ row index, λ column index) the rank of gC_τ^{-1} (mod p) is 1 or 0 according as to whether $S_\tau \not\equiv 0 \pmod p$, or $S_\tau \equiv 0 \pmod p$. The considerations in §17 show that for a block of the lowest kind, at least one of the elementary divisors of C_τ is $\geqq p^a$, $\alpha_\nu \geqq a$. It follows that gC_τ^{-1} has one elementary divisor 1, and we have

$$(55) \qquad S_\tau \not\equiv 0 \pmod p \quad (B_\tau \text{ block of the lowest kind}).$$

Since here gC_τ^{-1} has (mod p) one elementary divisor 1, C_τ has exactly one elementary divisor p^a. Consequently, the number of blocks of the lowest kind is equal to the number of elementary divisors p^a of C. This, in connection with the result of §16 yields theorem 1.

We add some remarks about the determination of the numbers S_τ for blocks of the lowest kind. From (12) it follows that $f_\kappa = g^{-1} \sum_\lambda \tilde{\gamma}_{\kappa\lambda} u_\lambda$. Combining this with (34), we obtain

$$f_\kappa = \frac{1}{g'} \sum \tilde{\gamma}_{\kappa\lambda} \frac{u_\lambda}{p^a} \equiv \frac{1}{g'} f_\kappa S_\tau \sum{}' f_\lambda \frac{u_\lambda}{p^a} \pmod p$$

where λ ranges over all values for which $\varphi^{(\lambda)}$ belongs to the block B_τ. If B_τ is of the lowest kind, we may assume $f_\kappa \not\equiv 0 \pmod p$. Hence

$$(56) \qquad g' \equiv S_\tau \sum{}' \frac{f_\lambda u_\lambda}{p^a} \pmod p \qquad (\varphi^{(\lambda)} \text{ in } B_\tau)$$

whence S_τ (mod p) can be obtained, if only the degrees of the characters are known. Using (12) and (5), we easily obtain

$$(57) \qquad \sum_i{}'' z_i^2 = \sum_i{}'' \sum_\lambda{}' (d_{i\lambda} f_\lambda)^2 = \sum_{\kappa,\lambda}{}' c_{\kappa\lambda} f_\kappa f_\lambda = \sum_\kappa{}' u_\kappa f_\kappa$$

where i ranges over those values for which $\zeta^{(i)}$ belongs to B_τ. Hence

$$(58) \qquad g' \equiv S_\tau \frac{\sum'' z_i^2}{p^a} \pmod{p}.$$

The numbers S_τ can also be determined in a different manner from the ordinary group characters $\zeta^{(i)}$ of $\mathfrak{G}$. We set

$$(59) \qquad V = (v_{ij}) = \left(\sum_{\nu=1}^{k} g_\nu \zeta_\nu^{(i)} \zeta_{\nu^*}^{(j)} \right).$$

We have then, making use of (21)

$$v_{ij} = \sum_{\nu=1}^{k} \sum_{\kappa=1}^{k} \sum_{\lambda=1}^{k} g_\nu d_{i\kappa} d_{j\lambda} \varphi_\nu^{(\kappa)} \varphi_{\nu^*}^{(\lambda)} = \sum_{\kappa,\lambda=1}^{k} d_{i\kappa} d_{j\lambda} \tilde{\gamma}_{\kappa\lambda}$$

$$(60) \qquad V = gDC^{-1}D' = gD(D'D)^{-1}D'.$$

Using (34), we obtain

$$v_{ij} \equiv \sum_{\kappa,\lambda} d_{i\kappa} d_{j\lambda} f_\kappa f_\lambda S_\tau \pmod{p}$$

if $\zeta^{(i)}$ and $\zeta^{(j)}$ both belong to B_τ. But $\sum_\kappa d_{i\kappa} f_\kappa = z_i$, according to (12), and hence

$$(61) \qquad \begin{aligned} v_{ij} &\equiv z_i z_j S_\tau \pmod{p} && (\zeta^{(i)} \text{ and } \zeta^{(j)} \text{ in } B_\tau) \\ v_{ij} &= 0 && (\zeta^{(i)} \text{ and } \zeta^{(j)} \text{ in different blocks}). \end{aligned}$$

IV. On the Multiplication of the Characters

22. Relations between the problems of determining the ordinary and the modular characters of $\mathfrak{G}$. For any group $\mathfrak{G}$ of order g, we have the two problems of finding the ordinary irreducible characters $\zeta^{(i)} (i = 1, 2, \cdots, n)$ and the modular irreducible characters $\varphi^{(\kappa)}$ $(k = 1, 2, \cdots, k)$ for a fixed prime p. We may assume that p divides g, since otherwise the two types of characters coincide. We ask now: (a) How much does knowing the ordinary characters help in the determination of the modular characters? (b) How much does knowing the modular characters help in the determination of the ordinary characters? It seems that in general we obtain some valuable information, but that in neither case the complete answer can be found. For instance, in the case of a p-group, the modular characters become trivial, since there is only the (1)-character, and this shows clearly that we cannot expect that the $\zeta_\nu^{(i)}$ are determined uniquely by the $\varphi_\nu^{(\kappa)}$. For both questions (a) and (b), it is of course of great importance to find the matrix D (cf. (13)).

If the $\varphi_\nu^{(\kappa)}$ are known, then (21) permits the determination of the matrix C, so that we also may find the characters $\eta^{(\kappa)}$ of the indecomposable constituents U_κ. For the determination of D, we have the formulas (5). In certain cases, these formulas are sufficient to find D, cf. the example of the group LF (2, p) in §31. But, in general, we must expect several possible solutions for D some of which may belong to other groups $H, K, \cdots$ which also have $(\varphi_\lambda^{(\kappa)})$ as their modular characters. There is, of course, only a finite number of possibilities for D. If D itself is known, then the values of the ordinary characters $\zeta_\nu^{(i)}$ for p-regular elements G of $\mathfrak{G}$ can be obtained from (10). There remains then the determination of the values of the characters for the other classes. Mod p, we can find these values from the values of the characters for the p-regular classes (cf. §§6, 18). Further, we obtain conditions from the orthogonality relations for group characters. Also the method of multiplying characters can be used with advantage. It may be mentioned that in many important cases it seems easier to find the modular characters than the ordinary characters. For instance, in the case of many simple groups, the analogy with semisimple continuous groups can be used in the modular theory.

Conversely, let us assume now that the ordinary characters $\zeta^{(i)}$ are known. It follows from (13) that $D = Z\Phi^{-1}$, which shows that each column of D is of the form

$$d_i = \sum_{\nu=1}^{k} \zeta_\nu^{(i)} \alpha_\nu, \tag{62}$$

that is, each column of D is a linear combination of those columns in the tables of ordinary characters which correspond to p-regular elements. The α_ν are the elements of a column of Φ^{-1} and, therefore, are not known, but we have some information about them. For instance, they are of form β/g' where β is an integer of the field generated by the $\zeta^{(i)}$ and $g = p^a g'$, $(g', p) = 1$. The d_i must be rational integers $\geqq 0$. Further restrictions are obtained from (13) and the form (28) of D, and from the fact that the determinant of $D'D$ is known. But these conditions are not enough to determine D uniquely, several cases will have to be considered. If D is known, then the modular characters are known. The equations (59) and (60) show that the matrix $D(D'D)^{-1}D'$ can be found if the $\zeta^{(i)}$ are known, but this does not provide any new information.

We may add some remarks in this connection. The condition (62) is, of course, equivalent to saying that each column of D is orthogonal to each column of Z which corresponds to a p-singular element of G. If a vector $x = (x_1, x_2, \cdots, x_n)$ is orthogonal to all these columns of Z, then x is a linear combination of the columns of D. Similarly, if $y = (y_1, y_2, \cdots, y_n)$ is orthogonal to all columns of D, then y is a linear combination of the columns of Z which correspond to p-singular elements of $\mathfrak{G}$.

If a relation

$$\sum_{i=1}^{n} \zeta^{(i)}(G)\beta_i = 0 \quad \text{(for all } p\text{-regular elements } G \text{ of } \mathfrak{G}) \tag{63}$$

where the β_i are independent of G, then $\beta = (\beta_1, \beta_2, \cdots, \beta_n)$ is a linear combination of the columns of Z which correspond to p-singular elements, and vice versa. Hence β is orthogonal to every column of D. We cut $(\beta_1, \beta_2, \cdots, \beta_n)$ into s pieces corresponding to the s blocks B_τ, and replace all the β_i by 0 except those which belong to a fixed piece. This modified vector β still is orthogonal to all the columns of D because of the form (28) of D. Hence the modified vector β still satisfies (63)

$$\sum{}' \zeta^{(i)}(G)\beta_i = 0$$

where i ranges over only those values for which $\zeta^{(i)}$ belongs to a fixed block B_τ.

THEOREM 6. *If a linear relation between the ordinary characters holds for all p-regular elements of $\mathfrak{G}$, then the relation remains true if we leave away all terms except those which contain the characters of a fixed block B_τ.*

23. The multiplication of characters. If F and H are two modular representations then $G \to F(G) \times H(G)$ gives a new representation $F \times H$. Since the characteristic roots of the Kronecker product $F(G) \times H(G)$ are obtained by multiplying each characteristic root of $F(G)$ into each characteristic root of $H(G)$, it follows easily that the character of $F \times H$ is obtained by multiplying the characters of F and H. Applying this to the irreducible characters $\varphi^{(\kappa)}$, $\varphi^{(\lambda)}$ we have that $\varphi^{(\kappa)} \cdot \varphi^{(\lambda)}$ is again a character of $\mathfrak{G}$, reducible or irreducible, and we obtain formulas

$$\varphi^{(\kappa)} \cdot \varphi^{(\lambda)} = \sum_\mu a_{\kappa\lambda\mu} \varphi^{(\mu)} \tag{64}$$

where the $a_{\kappa\lambda\mu}$ are rational integers, $a_{\kappa\lambda\mu} \geqq 0$. There is, of course some connection with the corresponding coefficients appearing in the multiplication of ordinary characters. If we have

$$\zeta^{(i)} \zeta^{(j)} = \sum_k b_{ijk} \zeta^{(k)} \tag{65}$$

then we express ζ by means of the φ (cf. (10)) and obtain

$$\zeta^{(i)} \zeta^{(j)} = \sum d_{i\kappa} d_{j\lambda} \varphi^{(\kappa)} \varphi^{(\lambda)} = \sum d_{i\kappa} d_{j\lambda} a_{\kappa\lambda\mu} \varphi^{(\mu)}$$

$$\zeta^{(i)} \zeta^{(j)} = \sum b_{ijk} \zeta^{(k)} = \sum b_{ijk} d_{k\mu} \varphi^{(\mu)}$$

$$\sum_{\kappa,\lambda} d_{i\kappa} d_{j\lambda} a_{\kappa\lambda\mu} = \sum_k b_{ijk} d_{k\mu}. \tag{66}$$

We derive some further relations for the $a_{\kappa\lambda\mu}$. Of course, they must satisfy the conditions for the constants of multiplication of a commutative algebra

$$a_{\kappa\lambda\mu} = a_{\lambda\kappa\mu}, \qquad \sum_\mu a_{\kappa\lambda\mu} a_{\mu\tau\sigma} = \sum_\mu a_{\kappa\mu\sigma} a_{\lambda\tau\mu}.$$

To each representation of $\mathfrak{G}$ there corresponds a contragredient representation. We denote the representation contragredient to F_κ by $F_{\kappa'}$. Then

$$\varphi_\nu^{(\kappa')} = \varphi_{\nu^*}^{(\kappa)} \tag{67}$$

where $\mathfrak{C}_{\nu^*}$ as in §7 denotes the class reciprocal to $\mathfrak{C}_\nu$. Here $1', 2', \cdots, k'$ and $1^*, 2^*, \cdots, k^*$ are permutations of period 2 of $1, 2, \cdots, k$. From (64) we obtain

$$a_{\kappa'\lambda'\mu'} = a_{\kappa\lambda\mu} \tag{68}$$

since the contragredient of $\varphi^{(\kappa)} \cdot \varphi^{(\lambda)}$ is $\varphi^{(\kappa')} \cdot \varphi^{(\lambda')}$. Further, the regular representation R is self-contragredient. Hence the contragredient of U_κ is an indecomposable constituent of R, and since its irreducible top constituent is $F_{\kappa'}$, we see that $U_{\kappa'}$ and U_κ are contragredient. This implies

$$c_{\kappa\lambda} = c_{\kappa'\lambda'}. \tag{69}$$

From (64) and the orthogonality relations (20) we obtain

$$g a_{\kappa\lambda\mu} = \sum_{\nu=1}^{k} g_\nu \varphi_\nu^{(\kappa)} \varphi_\nu^{(\lambda)} \eta_{\nu^*}^{(\mu)} \tag{70}$$

By multiplying the two left hand members through by $\gamma_{\mu\rho'} = \gamma_{\rho'\mu}$ and adding over μ we obtain

$$\sum_\mu g a_{\kappa\lambda\mu} \gamma_{\mu\rho'} = \sum_{\nu=1}^{k} g_\nu \varphi_\nu^{(\kappa)} \varphi_\nu^{(\lambda)} \varphi_{\nu^*}^{(\rho')}$$

or

$$\sum_\mu a_{\kappa\lambda\mu} \tilde{\gamma}_{\mu\rho'} = \sum_{\nu=1}^{k} g_\nu \varphi_\nu^{(\kappa)} \varphi_\nu^{(\lambda)} \varphi_\nu^{(\rho)}$$

which shows that the left side remains unchanged, when the three indices κ, λ, ρ are permuted. Thus

$$\sum a_{\kappa\lambda\mu} \tilde{\gamma}_{\mu\rho'} = \sum a_{\rho\lambda\mu} \tilde{\gamma}_{\mu\kappa'}. \tag{71}$$

In particular, for $\kappa = 1$, we have $a_{\kappa\lambda\mu} = \delta_{\lambda\mu}$, $\kappa' = 1$, and when we interchange ρ and ρ' we have

$$\tilde{\gamma}_{\lambda\rho} = \sum_\mu a_{\rho'\lambda\mu} \tilde{\gamma}_{\mu 1} \tag{72}$$

which shows that the whole matrix C^{-1} can be found if its first column and the constants of multiplication are known.

If $\varphi^{(\lambda)} \varphi^{(\rho')}$ does not contain a character of the block B_1, then the right side of (72) vanishes and $\tilde{\gamma}_{\lambda\rho} = 0$.

THEOREM 7.[29] *If the product of a character $\varphi^{(\lambda)}$ with the contragredient character $\varphi^{(\rho')}$ of $\varphi^{(\rho)}$ does not contain a character of the first block B_1, then the corresponding coefficient $\tilde{\gamma}_{\lambda\rho}$ of gC^{-1} vanishes.*

If the block B_τ of $\varphi^{(\lambda)}$ contains more than one modular character then because of the form (28) of C, $\tilde{\gamma}_{\lambda\rho} = 0$, cannot hold for all $\rho \neq \lambda$ such that $\varphi^{(\rho)}$ belongs

[29] This theorem is related to theorem 2 of R. Brauer, Math. Zeitschr. 41, 1936, p. 330.

to B_τ. Further, if $\varphi^{(\lambda)}$ and $\varphi^{(\rho)}$ belong to the same block, and if their degrees are not divisible by p, then $\tilde{\gamma}_{\lambda\rho} \neq 0$ according to (34) and (55). Hence we have the

COROLLARY. *If two characters $\varphi^{(\lambda)}$ and $\varphi^{(\rho)}$ belong to the same block and both have degrees prime to p, then $\varphi^{(\lambda)} \cdot \varphi^{(\rho)}$ contains a character of the first block.*

We prove two more formulas connecting the $c_{\kappa\lambda}$, the $a_{\kappa\lambda\mu}$, and the characters, and which deserve some interest. Using (9) we derive from (64)

$$\varphi_\nu^{(\kappa)} \eta_\nu^{(\mu)} = \sum_\lambda c_{\mu\lambda} \varphi_\nu^{(\kappa)} \varphi_\nu^{(\lambda)} = \sum_{\lambda,\rho} c_{\mu\lambda} a_{\kappa\lambda\rho} \varphi_\nu^{(\rho)}. \tag{73}$$

We first set $\mu = \kappa$, and add over κ. By (16) we find

$$\delta_{1\nu^*} g/g_\nu = \sum_{\kappa,\lambda,\rho} c_{\kappa\lambda} a_{\kappa\lambda\rho} \varphi_\nu^{(\rho)}.$$

Here, we multiply by $g_\nu \eta_{\nu^*}^{(\sigma)}$, add over ν, and use (20)

$$\sum{}' \eta_\nu^{(\sigma)} = \sum_{\kappa,\lambda} c_{\kappa\lambda} a_{\kappa\lambda\sigma} \tag{74}$$

where the sum on the left extends over those ν for which the class $\mathfrak{C}_\nu$ is self-reciprocal, $\mathfrak{C}_\nu = \mathfrak{C}_{\nu^*}$.

Secondly, we take $\mu = \kappa'$ in (73) and apply the same method. We thus obtain

$$\sum_{\nu=1}^{k} \eta_\nu^{(\sigma)} = \sum_{\kappa,\lambda} c_{\kappa\lambda} a_{\kappa'\lambda\sigma}. \tag{75}$$

It can easily be seen from (67) that the number of self-contragredient modular characters, $\varphi^{(\lambda)} = \varphi^{(\lambda')}$ is equal to the number of self-reciprocal p-regular classes, $\mathfrak{C}_\nu = \mathfrak{C}_{\nu^*}$ $(\nu \leq k)$.

24. Upper and lower bounds for the degrees of the indecomposable constituents U_κ. The product $\eta^{(\mu)} \cdot \varphi^{(\lambda')}$ can be expressed as a linear combination of the $\varphi^{(\rho)}$ (cf. (73)), and also, using (11), as a linear combination of the $\eta^{(\kappa)}$

$$\eta^{(\mu)} \cdot \varphi^{(\lambda')} = \sum \tilde{a}_{\kappa\lambda\mu} \eta^{(\kappa)}.$$

Here it is neither obvious that the coefficients are integers, nor that they are $\geqq 0$, but both these facts will follow from (73). Using (20) we find

$$g\tilde{a}_{\kappa\lambda\mu} = \sum_\nu g_\nu \eta_\nu^{(\mu)} \varphi_\nu^{(\lambda')} \varphi_\nu^{(\kappa')}$$

and comparing this with (70), and taking (68) into account, we obtain $\tilde{a}_{\kappa\lambda\mu} = a_{\kappa'\lambda'\mu'} = a_{\kappa\lambda\mu}$. Hence[30]

$$\eta^{(\mu)} \cdot \varphi^{(\lambda')} = \sum a_{\kappa\lambda\mu} \eta^{(\kappa)}. \tag{76}$$

[30] This formula seems to indicate that $U_\mu \times F_{\lambda'}$ splits completely into $U_1, U_2, \cdots, U_k$ where U_κ appears $a_{\kappa\lambda\mu}$ times, but we have not been able to prove this.

In particular, we have $a_{\kappa\kappa'1} \geqq 1$, since $F_\kappa \times F_{\kappa'}$ contains the 1-representation. Hence $\eta^{(\kappa)}$ will appear in $\eta^{(1)} \cdot \varphi^{(\kappa)}$. On comparing the degrees, we find $u_1 f_\kappa \geqq u_\kappa$. On the other hand $a_{1\kappa\kappa} = 1$ and hence $\eta^{(1)}$ will appear in $\eta^{(\kappa)} \cdot \varphi^{(\kappa')}$. Consequently $u_1 \leqq f_\kappa u_\kappa$.

THEOREM 8. *For the degrees f_κ of the F_κ and u_κ of the U_κ there hold the inequalities*

$$u_1 f_\kappa \geqq u_\kappa \geqq u_1/f_\kappa. \tag{77}$$

Since $u_\kappa = \sum_\lambda c_{\kappa\lambda} f_\lambda$ (cf. (12)), it follows from (77) that

$$u_\kappa \geqq c_{\kappa\kappa} f_\kappa + c_{\kappa\lambda} f_\lambda \qquad (\kappa \neq \lambda)$$

$$c_{\kappa\lambda} \leqq (u_1 - c_{\kappa\kappa}) \frac{f_\kappa}{f_\lambda} \quad \text{for} \quad \kappa \neq \lambda. \tag{78}$$

In particular, $c_{\kappa\lambda} < u_1$ since we may assume $f_\lambda \geqq f_\kappa$, further $c_{\kappa\kappa} \leqq u_1$.

On multiplying (77) by f_κ and adding, we find

$$u_1 \sum_{\kappa=1}^{k} f_\kappa^2 \geqq \sum_{\kappa=1}^{k} u_\kappa f_\kappa \geqq k u_1$$

The middle term here is g, as follows from (16). If the radical of the modular group ring Γ has the order m, then $\sum f_\kappa^2 = g - m$. Hence

$$u_1(g - m) \geqq g \geqq k u_1$$

$$g/k \geqq u_1 \geqq g/(g - m). \tag{79}$$

The multiplication of characters is used to obtain new characters if some characters have already been found. It is often convenient to determine the $\eta^{(\kappa)}$ at the same time with the $\varphi^{(\kappa)}$. Here formula (76) can be used. Formulas (77) and (79) can sometimes be used, if we want to show that a character η, which we have obtained, is an $\eta^{(\kappa)}$ and not a sum of several such $\eta^{(\kappa)}$.

V. Relations Between the Characters of a Group $\mathfrak{G}$ and Those of a Subgroup $\mathfrak{H}$

25. The induced character. The second important method of Frobenius for the construction of characters assumes that the character χ of a representation V of a subgroup $\mathfrak{H}$ of $\mathfrak{G}$ is known. This representation "induces" a representation V^* of $\mathfrak{G}$ whose character χ^* can be obtained. The method of forming V^* remains valid in case we start with a modular representation, and so does the formula for χ^*, but this last formula requires a somewhat different proof here, due to the modified definition of the character of a representation.

Let h be the order of $\mathfrak{H}$, and let Q_μ ($\mu = 1, 2, \cdots, m$; $m = g/h$) be a complete residue system of $\mathfrak{G}$ (mod $\mathfrak{H}$)

$$\mathfrak{G} = \mathfrak{H}Q_1 + \mathfrak{H}Q_2 + \cdots + \mathfrak{H}Q_m.$$

We set $V(G) = 0$ if G does not belong to $\mathfrak{H}$ so that $F(G)$ is defined for all elements of $\mathfrak{G}$, and define

$$V^*(G) = (V(Q_\kappa G Q_\lambda^{-1})) \qquad (\kappa \text{ row index, } \lambda \text{ column index}). \tag{80}$$

It is easily seen that this is a representation V^* of $\mathfrak{G}$ of degree $tm = tg/h$ where t is the degree of V.

We shall determine the character of V^*. Let G be an arbitrary element of $\mathfrak{G}$. The element $Q_\mu G$ belongs to some residue class $\mathfrak{H}Q_{\rho_G(\mu)}$ and the permutations $P_G: \mu \to \rho_G(\mu)$ $(\mu = 1, 2, \cdots, m)$ form a representation $P_\mathfrak{G}$ of $\mathfrak{G}$.

We split P_G into cycles. The length of each cycle is a divisor of the order of G. If G is p-regular, then the length of each cycle is prime to p. Let, for instance, $(1, 2, \cdots, \alpha)$ be the first cycle of P_G for a p-regular element G. Then $V^*(G)$ breaks up completely into the matrix

$$W = \begin{pmatrix} 0 & V(Q_1GQ_2^{-1}) & 0 & \cdots & 0 \\ 0 & 0 & V(Q_2GQ_3^{-1}) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & V(Q_{\alpha-1}GQ_\alpha^{-1}) \\ V(Q_\alpha GQ_1^{-1}) & 0 & 0 & \cdots & 0 \end{pmatrix}$$

and analagous matrices, corresponding to the other cycles of P_G. We have to determine the characteristic equation of W. We set $W_\nu = V(Q_\nu GQ_{\nu+1}^{-1})$ $(\nu = 1, 2, \cdots, \alpha - 1)$, $W_\alpha = V(Q_\alpha GQ_1^{-1})$. We multiply the νth column of (81) on its right side by $W_{\nu-1}^{-1}W_{\nu-2}^{-1} \cdots W_1^{-1}$, and the νth row on its left side by $W_1W_2 \cdots W_{\nu-1}$ $(\nu = 2, 3, \cdots, \alpha)$. Then $W_1, W_2, \cdots, W_{\alpha-1}$ in (81) are replaced by the unit matrix, whereas we have

$$\begin{aligned} W_1W_2 \cdots W_\alpha &= V(Q_1GQ_2^{-1})V(Q_2GQ_3^{-1}) \cdots V(Q_{\alpha-1}GQ_\alpha^{-1})V(Q_\alpha GQ_1^{-1}) \\ &= V(Q_1G^\alpha Q_1^{-1}) \end{aligned}$$

at the place of W_α in the last row, first column. Obviously, our changing of W amounts to a similarity transformation so that the characteristic polynomial remains unaltered. For the new form of W, the characteristic polynomial can be easily computed, and we find $f(x^\alpha)$ where $f(x)$ denotes the characteristic polynomial of $Q_1G^\alpha Q_1^{-1} = H$. As product of the elements $Q_\nu GQ_{\nu+1}^{-1}$ $(\nu = 1, 2, \cdots, \alpha - 1)$ and $Q_\alpha GQ_1^{-1}$, this element H belongs to $\mathfrak{H}$. It is p-regular, since it is conjugate to G^α in $\mathfrak{G}$. Hence we obtain the characteristic roots of W by taking all the α^{th} roots of the characteristic roots of $V(H)$. This is still valid if we replace the characteristic roots (which lie in the modular field $\bar{K}$) by the corresponding complex roots of unity. Since α and the order of H are prime to p, no difficulty arises. It follows easily that if $\alpha > 1$ then the sum of these complex roots of unity must vanish. If $\alpha = 1$, then $W = V(Q_1GQ_1^{-1})$, and the sum in this case is the character $\chi(Q_1GQ_1^{-1})$. Dealing with the other cycles of P_G in the same manner, we obtain $\sum \chi(Q_\mu GQ_\mu^{-1})$ for the value of the character $\chi^*(G)$

of V^*. Here μ ranges over all values for which $Q_\mu G Q_\mu^{-1}$ lies in $\mathfrak{H}$. If we set $\chi(G) = 0$ for elements G outside of $\mathfrak{H}$ we may write

$$\chi^*(G) = \sum_{\mu=1}^{m} \chi(Q_\mu G Q_\mu^{-1}). \tag{82}$$

This is exactly the formula of Frobenius, only the trace argument by which it is ordinarily derived from (80) could not be used because of our modification in the definition of the modular characters (§6).

The splitting of the regular representation of $\mathfrak{H}$ into its indecomposable constituents corresponds to a decomposition of the group ring of $\mathfrak{H}$ into a direct sum of right ideals $\mathfrak{f}_\rho$. The products $h_\rho Q_\mu$ ($\mu = 1, 2, \cdots, m$) of Q_μ with the elements of $\mathfrak{f}_\rho$ generate a right ideal $\mathfrak{f}_\rho^*$ of the group ring Γ, and Γ is the direct sum of the $\mathfrak{f}_\rho^*$. If $\mathfrak{f}_\rho$ corresponds to the representation V_ρ of $\mathfrak{H}$ then it is known that $\mathfrak{f}_\rho^*$ corresponds to the representation V_ρ^* of $\mathfrak{G}$. It follows that the regular representation of $\mathfrak{G}$ breaks up completely into the V_ρ^*, corresponding to the different values of ρ. Each V_ρ^* itself consists of one or several of the indecomposable constituents U_κ.

We use for $\mathfrak{H}$ the same notations as for $\mathfrak{G}$ but with a $\sim$ sign, so $\tilde{F}_\kappa^*$ are the modular irreducible representations of $\mathfrak{H}$ etc. It follows now that $\tilde{U}_\kappa^*$ breaks up completely into some of the U_λ.

26. The formulas of Nakayama. We now can prove easily that we have formulas[31]

$$\begin{cases} \tilde{\eta}^{(\kappa)*} = \sum_\lambda \alpha_{\kappa\lambda} \eta^{(\lambda)} & \text{(for } p\text{-regular elements of } \mathfrak{G}) \\ \varphi^{(\lambda)} = \sum_\kappa \alpha_{\kappa\lambda} \tilde{\varphi}^{(\kappa)} & \text{(for } p\text{-regular elements of } \mathfrak{H}) \end{cases} \tag{83}$$

where the $\alpha_{\kappa\lambda}$ are rational integers, $\alpha_{\kappa\lambda} \geqq 0$. Obviously, the only point which requires a proof is that the same coefficients $\alpha_{\kappa\lambda}$ appear in both formulas. But from the orthogonality relations for the modular group characters (cf. (16), (20)) it follows for the coefficients $\alpha_{\kappa\lambda}$ of the first equation (83) that

$$g\alpha_{\kappa\lambda} = \sum \tilde{\eta}^{(\kappa)*}(G)\varphi^{(\lambda)}(G^{-1})$$

where the sum extends over all the p-regular elements G of $\mathfrak{G}$. Using (82) we obtain

$$g\alpha_{\kappa\lambda} = \sum_G \sum_{\mu=1}^{m} \tilde{\eta}^{(\kappa)}(Q_\mu G Q_\mu^{-1})\varphi^{(\lambda)}(G^{-1})$$

and after a simple rearrangement of the terms

$$h\alpha_{\kappa\lambda} = \sum \tilde{\eta}^{(\kappa)}(H)\varphi^{(\lambda)}(H^{-1})$$

[31] The formulas (83) and (85) are equivalent to those given in theorem 9 of T. Nakayama, Ann. of Math. 39, 1938, p. 361.

where H ranges over all the p-regular elements of $\mathfrak{H}$. This shows that $\alpha_{\kappa\lambda}$ is exactly the coefficient appearing in the second formula (83). This is exactly the method by which it is shown for ordinary characters that we have formulas

$$(84)\qquad \begin{cases} \tilde{\zeta}^{(i)*} = \sum_j l_{ij}\zeta^{(j)} & \text{(for elements of } \mathfrak{G}) \\ \zeta^{(j)} = \sum_i l_{ij}\tilde{\zeta}^{(i)} & \text{(for elements of } \mathfrak{H}) \end{cases}$$

where the l_{ij} are rational integers, $l_{ij} \geqq 0$.

Finally, we have formulas

$$(85)\qquad \begin{cases} \tilde{\varphi}^{(\kappa)*} = \sum_\lambda \beta_{\kappa\lambda}\varphi^{(\lambda)} & \text{(for } p\text{-regular elements of } \mathfrak{G}) \\ \eta^{(\lambda)} = \sum_\kappa \beta_{\kappa\lambda}\tilde{\eta}^{(\kappa)} & \text{(for } p\text{-regular elements of } \mathfrak{H}) \end{cases}$$

with rational integers $\beta_{\kappa\lambda} \geqq 0$, as can be shown in the same manner, or also be derived from (83).

We set $A = (\alpha_{\kappa\lambda})$, $B = (\beta_{\kappa\lambda})$, $L = (l_{ij})$. On comparing (84), (85) and (83) we obtain

$$\tilde{D}B = LD, \qquad DA' = L'\tilde{D}$$
$$AC = \tilde{C}B$$

where $\tilde{D}$, $\tilde{C}$ have the same significance for $\mathfrak{H}$ as D, C have for $\mathfrak{G}$.

The second formula (83) shows that $\alpha_{11} = 1$, $\alpha_{\kappa 1} = 0$ for $\kappa \neq 1$ if the index 1 always refers to the 1-representation. Hence $\eta^{(1)}$ appears in the character $\tilde{\eta}^{(1)*}$ of degree $\tilde{u}_1 \cdot g/h$. Hence

$$(86)\qquad u_1 \leqq \tilde{u}_1 g/h.$$

In particular, if $\mathfrak{H}$ has an order prime to p, we have $\tilde{u}_1 = 1$, and obtain

THEOREM 9. *The degree of the indecomposable constituent of the regular representation of $\mathfrak{G}$ which corresponds to the* 1*-representation, is at most equal to the index of the maximal subgroup $\mathfrak{H}$ of an order prime to p.*

In particular, if $\mathfrak{G}$ has a subgroup $\mathfrak{H}$ of index p^a, then $u_1 \leqq p^a$, and since u_1 is divisible by p^a, we have $u_1 = p^a$. It follows that in this case U_1 is identical with the representation of $\mathfrak{G}$ by permutations which corresponds to the subgroup $\mathfrak{H}$ of index p^a (cf. the remark at the end of §25).

In the general case, it follows from (83) that every $\eta^{(\lambda)}$ appears in at least one $\tilde{\eta}^{(\kappa)*}$. Hence

$$(87)\qquad u_\lambda \leqq (g/h)\cdot \max(\tilde{u}_\kappa).$$

If $\mathfrak{H}$ again has an order prime to p, then $\tilde{u}_\kappa = \tilde{z}_\kappa = \tilde{f}_\kappa$, and we have $u_\lambda \leqq (g/h)\max(\tilde{f}_\kappa)$.

27. On the converse of theorem 1. We consider now the case that $\mathfrak{H}$ is the Sylow subgroup of order p^a of $\mathfrak{G}$. The character $\tilde{\varphi}^{(1)*}$ has here the value

$g' = g/p^a$ for the unit element and the value 0 for all the other p-regular elements. Hence

$$\sum_{\nu=1}^{k} g_\nu \tilde{\varphi}_\nu^{(1)*} \eta_{\nu*}^{(\lambda)} = g' u_\lambda = g \frac{u_\lambda}{p^a}.$$

Using the orthogonality relations (16), we obtain u_λ/p^a as multiplicity of $\varphi^{(\lambda)}$ in $\tilde{\varphi}^{(1)*}$

(88) $$\tilde{\varphi}^{(1)*} = \sum_{\lambda=1}^{k} \frac{u_\lambda}{p^a} \varphi^{(\lambda)}.[32]$$

From (84) we have

(89) $$\tilde{\zeta}^{(1)*} = \sum_i l_i \zeta^{(i)} \qquad (l_i = l_{1i}).$$

For p-regular elements, these two characters are identical. By (12) and (10) we have $u_\lambda = \sum d_{i\lambda} z_i$, $\zeta^{(i)} = \sum d_{i\lambda} \varphi^{(\lambda)}$, and thus obtain

$$\sum_\lambda \sum_i \frac{d_{i\lambda} z_i}{p^a} \varphi^{(\lambda)} = \sum \sum l_i d_{i\lambda} \varphi^{(\lambda)}.$$

for all p-regular elements of $\mathfrak{G}$. Hence, on comparing the coefficients of $\varphi^{(\lambda)}$

(90) $$\sum_i \left(\frac{z_i}{p^a} - l_i\right) d_{i\lambda} = 0 \qquad (\lambda = 1, 2, \cdots, k).$$

Assume now that we have a block B_τ which contains the same number of ordinary and modular representations, $x_\tau = y_\tau$. We choose the index λ in (90) such that $\varphi^{(\lambda)}$ belongs to B_τ. Then it is sufficient to let i range over those values for which $\zeta^{(i)}$ belongs to B_τ, since for other values of i, $d_{i\lambda} = 0$. We may consider (90) as a system of y_τ linear homogeneous equations for the $x_\tau = y_\tau$ quantities $z_i/p^a - l_i$. Since the determinant is D_τ' and has the rank y_τ, all the $(z_i/p^a) - l_i$ vanish. But l_i is an integer, hence $z_i \equiv 0 \pmod{p^a}$. This shows again, that if $x_\tau = y_\tau$ then B_τ is a block of highest kind (converse of the middle part of theorem 1, cf. §14).

Using (84), we see that l_i in (89) is also the coefficient with which $\tilde{\zeta}^{(1)}$ appears in $\zeta^{(i)}$. Hence

(91) $$l_i = \frac{1}{p^a} \sum \zeta^{(i)}(H)$$

where H ranges over all the elements of the Sylow group $\mathfrak{H}$ of order p^a. The first term here is z_i/p^a. Combining this formula with the results of §22, we easily can obtain (90) again.

The formula (91) shows that if $\zeta^{(i)}(H)$ vanishes for all elements of an order

[32] This formula shows Dickson's theorem, $p^a \mid u_\lambda$ (cf. footnote 6), and this is essentially the way by which Dickson proved his result.

p^μ with $\mu > 0$, then $l_i = z_i/p^a$. Hence in this case $\zeta^{(i)}$ must be a character of the highest kind.

THEOREM 10. *If an irreducible character $\zeta^{(i)}$ vanishes for all elements of an order p^μ, $\mu > 0$, then $\zeta^{(i)}$ is a character of the highest kind.*

This is a converse to the last part of theorem 1. It even would be sufficient to assume that all the $\zeta^{(i)}(H)$ are divisible by p^a, if H is an element of order p^μ, $\mu > 0$. If all these $\zeta^{(i)}(H)$ are divisible by p^a, then $p^a \mid z_i$. This result can easily be improved when we take into account the multiplicity with which the terms $\zeta^{(i)}(H)$ appear on the right side of (91).

28. An upper and a lower bound for c_{11}. We now consider two arbitrary subgroups $\mathfrak{H}$ and $\mathfrak{J}$ of $\mathfrak{G}$ of orders h and j. Let $\alpha(G)$ and $\beta(G)$ be the ordinary characters of $\mathfrak{G}$, induced by the 1-representations of $\mathfrak{H}$ and $\mathfrak{J}$ respectively. From (82) we obtain easily that $\alpha(G)h$ is the number of elements M in $\mathfrak{G}$ for which G lies in $M^{-1}\mathfrak{H}M$, and similarly $\beta(G)j$ is the number of elements N for which G lies in $N^{-1}\mathfrak{J}N$. Then G will lie in $hj\alpha(G)\beta(G)$ of the intersections $\mathfrak{D}_{M,N} = [M^{-1}\mathfrak{H}M,\ N^{-1}\mathfrak{J}N]$.

If $\mathfrak{D}_{M,N}$ has the order $t_{M,N}$ then

$$(92) \qquad hj \sum_G \alpha(G)\beta(G) = \sum_{M,N} t_{M,N} = \sum_{M,N} t_{MN^{-1},1} = g \sum t_{M,1}.$$

On the other hand, we split $\mathfrak{G}$ into residue classes modd $\mathfrak{H}$ and $\mathfrak{J}$.

$$(93) \qquad \mathfrak{G} = \sum_{\nu=1}^{r} \mathfrak{H}R_\nu\mathfrak{J}.$$

The number of elements in $\mathfrak{H}R_\nu\mathfrak{J}$ is equal to $hj/t_{M,1}$ where M is any element of $\mathfrak{H}R_\nu\mathfrak{J}$. Hence, if M ranges over the elements of $\mathfrak{H}R_\nu\mathfrak{J}$, $\sum' t_{M,1} = hj$, and from (92) it follows that

$$(94) \qquad \begin{aligned} hj \sum_G \alpha(G)\beta(G) &= g\cdot rhj \\ \sum_G \alpha(G)\beta(G) &= g\cdot r \end{aligned}$$

where r is the number of residue classes of $\mathfrak{G}$ (modd $\mathfrak{H}$, $\mathfrak{J}$).

We assume now that $\mathfrak{H}$ and $\mathfrak{J}$ have orders prime to p. We may restrict the summation on the left hand of (94) to p-regular elements since for the other elements $\alpha(G) = 0$, and may consider $\alpha(G)$ and $\beta(G)$ as the modular characters of $\mathfrak{G}$, induced by the 1-representations of $\mathfrak{H}$ and $\mathfrak{J}$. We may set

$$\alpha(G) = \sum a_\kappa \eta^{(\kappa)}(G), \qquad \beta(G) = \sum b_\lambda \eta^{(\lambda)}(G)$$

where according to (83), a_κ and b_λ are rational integers $\geqq 0$, and $a_1 = 1$, $b_1 = 1$.

$$\alpha(G) = \sum_{\kappa,\lambda} a_\kappa c_{\kappa\lambda} \varphi^{(\lambda)}(G).$$

It is easily verified by use of (82) that $\beta(G) = \beta(G^{-1})$ and so combining (94) and the orthogonality relations (20), we obtain

$$\sum_{\kappa,\lambda=1}^{k} a_\kappa c_{\kappa\lambda} b_\lambda = r. \tag{95}$$

In particular,

$$c_{11} \leqq r. \tag{96}$$

THEOREM 11. *If $\mathfrak{H}$ and $\mathfrak{J}$ are two subgroups of $\mathfrak{G}$, whose orders are prime to p, then the first Cartan invariant c_{11} is at most equal to the number of residue classes of $\mathfrak{G}$* (mod $\mathfrak{H}$, $\mathfrak{J}$).

If, for instance, $\mathfrak{G}$ is a doubly transitive permutation group of order p^a, then we may take $\mathfrak{H} = \mathfrak{J}$ as subgroup of index p^a. Here $r = 2$, and hence $c_{11} = 2$.

Since C is the matrix of a positive definite quadratic form, the coefficient in the first row and first column of C^{-1} is at least $\frac{1}{c_{11}}$, and the equality sign is possible only if $c_{1s} = 0$ for all $s > 1$, i.e. if the modular 1-representation forms a block of its own. In this exceptional case, we have $C_1 = (c_{11})$, $c_{11} = p^a$, since p^a is the only elementary divisor of C_1. Then $\mathfrak{G}$ contains a normal subgroup $\mathfrak{H}$ of index p^a, (cf. §29 below).

In any case, we find from $gC^{-1} = (\gamma_{\kappa\lambda})$ and (35)

$$N = \tilde{\gamma}_{11} \geqq g/c_{11}$$

$$c_{11} \geqq g/N. \tag{97}$$

THEOREM 12. *The first Cartan invariant c_{11} is at least equal to g/N where N is the number of elements of an order prime to p in $\mathfrak{G}$. The equality sign holds only if these elements form a normal subgroup, necessarily of index p^a.*

From (96) and (97) it follows that $rN > g$, except for the case that $\mathfrak{G}$ contains a normal subgroup of index p^a. If, for instance, g is divisible by three distinct primes $g = p^a p_1^{a'} p_2^{a''}$, then we can take for $\mathfrak{H}$ and $\mathfrak{J}$ the Sylow-groups of orders $p_1^{a'}$, and $p_2^{a''}$ and have $r = p^a$. Hence $N > g/p^a$, except when $\mathfrak{G}$ contains a normal subgroup of index p^a.[33]

VI. SPECIAL CASES AND EXAMPLES

29. Special cases. We first consider the case that $\mathfrak{G}$ is a direct product, $\mathfrak{G} = \mathfrak{A} \times \mathfrak{B}$. If $A \to F(A)$ is a representation of $\mathfrak{A}$, and $B \to K(B)$ is a representation of $\mathfrak{B}$, then $A \times B \to F(A) \times K(B)$ (Kronecker product) is a representation of $\mathfrak{A} \times \mathfrak{B}$. This representation $F \times K$ is irreducible, if F and K are irreducible, and conversely, every irreducible representation of $\mathfrak{A} \times \mathfrak{B}$ is of

[33] This is a very special case of an unproved conjecture of Frobenius which states that when there are exactly r elements X an order dividing r in a group of these elements form a subgroup.

this form.[34] This implies that the D-matrix of $\mathfrak{G}$ is the direct product of the D-matrices of $\mathfrak{A}$ and of $\mathfrak{B}$, and that the C-matrix of $\mathfrak{G}$ is the product of the C-matrices of $\mathfrak{A}$ and of $\mathfrak{B}$.

We next consider the case that $\mathfrak{G}$ contains a normal subgroup $\mathfrak{H}$ of index $p^a(g = p^a g'$ with $(g', p) = 1)$. Since $\eta^{(1)} = \tilde{\varphi}^{(1)*}$ (cf. §26). we see that $\eta^{(1)}$ has the value p^a for every p-regular element. It follows that $\eta^{(1)} = p^a\varphi^{(1)}$, $C_1 = (p^a)$, and $\varphi^{(1)}$ is the only irreducible modular character in the first block B_1. Conversely, let $\mathfrak{G}$ be a group for which the first block B_1 contains only one irreducible modular character. Denote by $\mathfrak{H}$ the normal subgroup whose elements are represented by the unit matrix I in each ordinary irreducible representation Z_i of the first block. Then Z_i is a representation of $\mathfrak{G}/\mathfrak{H}$, and the index g/h is at least equal to the sum of the squares of the degrees z_i of these Z_i. This sum is equal to $u_1 f_1 \geqq p^a$, hence $g/h \geqq p^a$. On the other hand, each p-regular element of $\mathfrak{G}$ is represented in Z_i by a matrix whose characteristic roots are all 1, and which then is equal to I. This shows that h is divisible by every prime power dividing g/p^a. Hence $h \geqq g/p^a$, so $h = g/p^a$. We see that $\mathfrak{G}$ contains a normal subgroup of index p^a, if and only if, the first block contains only one irreducible modular character.

Let us assume now that $\mathfrak{G}$ contains a normal subgroup $\mathfrak{P}$ of order p^a, $(g = p^a g', (g', p) = 1)$. Every representation of $\mathfrak{G}/\mathfrak{P}$ defines a representation of $\mathfrak{G}$. In particular, the regular representation of $\mathfrak{G}/\mathfrak{P}$ has the character χ, $\chi(1) = g/p^a$, $\chi(G) = 0$ for a p-regular element $G \neq 1$. This is the character $\tilde{\varphi}^{(1)*}$ of §27. From (88), it follows that this character contains $\varphi^{(\lambda)}$ exactly u_λ/p^a times. In particular, $\varphi^{(\lambda)}$ must represent the elements of $\mathfrak{P}$ by the unit matrix.[35] Further, since the regular representation of $\mathfrak{G}/\mathfrak{P}$ contains each of its irreducible constituents $\varphi^{(\lambda)}$ exactly f_λ times, we have $f_\lambda = u_\lambda/p^a$. The degree f_λ is prime to p, since $\varphi^{(\lambda)}$ is a representation of $\mathfrak{G}/\mathfrak{P}$ of order g', so each block is of the lowest kind. We do not know whether the converse is true.

Finally, let $\mathfrak{G}$ be a group in which every p-regular element commutes with every element of a p-Sylow group. Then we have k blocks of the lowest kind. Each of them can contain only one modular irreducible constituent, and C necessarily has the form

$$C = \begin{pmatrix} p^a & & & \\ & p^a & & 0 \\ & & \ddots & \\ & 0 & & p^a \end{pmatrix}.$$

The converse is also true.

In the same manner as for ordinary characters, it follows that every linear character of any group $\mathfrak{G}$ is actually a character of $\mathfrak{G}/\mathfrak{K}$ where $\mathfrak{K}$ denotes the

[34] The first part follows easily from Burnside's theorem, Proc. London Math. Soc. (2) 3, 1905, p. 430; the second from A. H. Clifford's theorem, Ann. of Math. 38, 1937, p. 533.

[35] This follows from the fact that F_λ appears as a constituent in a representation for which this is true.

commutator group of $\mathfrak{G}$. Hence, the number of linear modular characters of $\mathfrak{G}$ is equal to the largest factor of the index of the commutator group which is prime to p. For a linear character $\varphi^{(\lambda)}$, the formula (77) shows that $u_\lambda = u_1$. This can also be shown directly, as U_λ may be expressed as the direct product of U_1 and $\varphi^{(\lambda)}$ (cf. (76), (64)). The linear character $\varphi^{(\lambda)}$ will belong to the block B_1, if and only if $\varphi^{(\lambda)} = 1$ for every p-regular element G which commutes with every element of a p-Sylow group. The block B_τ of $\varphi^{(\lambda)}$ is obtained from the block B_1 by multiplication with $\varphi^{(\lambda)}$.

30. The groups $GLH(2, p^a)$, $SLH(2, p^a)$, and $LF(2, p^a)$. As first examples we treat the group $SLH(2, p^a)$ of all matrices $\begin{pmatrix} m_{11}, & m_{12} \\ m_{21}, & m_{22} \end{pmatrix} = G$ of determinant 1 with coefficients in the Galois field $GF(p^a)$ with $p^a = q$ elements. The rth power matrices $G^{(r)}$ form a representation $\mathfrak{G}^{(r)}$ for any fixed $r = 0, 1, 2, \cdots$. Further, if $\mathfrak{H}$ is a modular representation with coefficients in $GF(q)$, we obtain a new representation by applying an automorphism θ of $GF(q)$ to all coefficients; we denote this representation by $\mathfrak{H}^\theta$. Let θ_ν now be the automorphism $\alpha \to \alpha^{p^\nu}$ of $GF(q)$ $(\nu = 1, 2, \cdots, a - 1)$. We form the representation

$$(98) \qquad H(r_0, r_1, \cdots, r_{a-1}) = \mathfrak{G}^{(r_0)} \times \mathfrak{G}^{(r_1)\theta_1} \times \cdots \times \mathfrak{G}^{(r_{a-1})\theta_{a-1}}$$

$(r_\nu = 0, 1, 2, \cdots, p - 1)$. We thus obtain p^a modular representations, and state that these are all the irreducible representations. In order to prove this, we first notice that there are exactly p^a classes of p-regular conjugate elements in $\mathfrak{G}$ since each such class corresponds in a $(1 - 1)$ manner to a polynomial $x^2 - tr(G)\, x + 1$, the characteristic polynomial of its elements, and $tr(G)$ can be any element of $GF(q)$. Secondly, we prove that the representations (98) are irreducible. Let x_ν, y_ν undergo the transformation G^{θ_ν}, $(\nu = 0, 1, 2, \cdots, a - 1)$. Then (98) belongs to the vector-module $\mathfrak{B}$ of all polynomials in x_0, $y_0, \cdots, x_{a-1}, y_{a-1}$ which are homogeneous of degree r_ν in x_ν, y_ν, $(\nu = 0, 1, \cdots, a - 1)$.[36] We have to show that $\mathfrak{B}$ is irreducible, when the elements of $\mathfrak{G}$ are taken as operators. If F now is any element of $\mathfrak{B}$, the module $\mathfrak{M}(F)$ generated by F will contain all the polynomials F_t which are obtained from F by applying $x \to x + ty$, $y \to y$ for any t in $GF(q)$. Then $x_\nu \to x_\nu + t^{p^{\nu}-1} y_\nu$, $y_\nu \to y_\nu$. Obviously, F_t is of the form

$$F_t = H_0 + tH_1 + \cdots + t^{p^a-1} H_{p^a-1},$$

where H_μ depends on x_ν, y_ν. We now take for t the q different elements of $GF(q)$. It follows easily that each H_μ is a linear combination of the F_t, and hence lies in $\mathfrak{M}(F)$. The last H_μ which is not zero obviously is a single power product $Ay_0^{r_0}y_1^{r_1} \cdots y_{a-1}^{r_{a-1}}$, and hence $y_0^{r_0}y_1^{r_1} \cdots y_{a-1}^{r_{a-1}}$ lies in $\mathfrak{M}(F)$. We replace F now by this polynomial, apply the transformation $x \to x$, $y \to tx + y$, and use the same argument. We thus see that every power product of $x_0, y_0, \cdots, x_{a-1}$, y_{a-1} of the correct degrees lies in $\mathfrak{M}(F)$. Hence $\mathfrak{M}(F) = \mathfrak{B}$, i.e. (98) is irreducible.

[36] The coefficients of these polynomials can be taken from any extension field of $GF(q)$.

Finally, the representations (98) are all distinct. In order to show that, assume

$$H(r_0, r_1, \cdots, r_{a-1}) = H(r'_0, r'_1, \cdots, r'_{a-1}) \tag{99}$$

$0 \leqq r_i \leqq p-1$, $0 \leqq r'_i \leqq p-1$, and $r_i = r'_i$ does not hold for all i. We arrange the $H(r_0, r_1, \cdots, r_{a-1})$ in lexicographical order by taking $H(r_0, r_1, \cdots, r_{a-1})$ as lower than $H(s_0, s_1, \cdots, s_{a-1})$ when the first difference $s_0 - r_0$, $s_1 - r_1$, $\cdots$ $s_{a-1} - r_{a-1}$ which does not vanish, has a positive value. We may assume that $H(r_0, r_1, \cdots, r_{a-1})$ is the lowest representation (98) which is similar to another of these representations. Certainly not all the r can be equal to $p - 1$. But if all the r'_ν were equal to $p - 1$ then the right side in (99) would have the maximum degree p^a which would imply that all the $r_i = p - 1 = r'_i$. This case is, therefore, also excluded.

Assume $r_0 = \cdots = r_{i-1} = 0$, $r_i \neq 0$, $i \geqq 0$. We multiply (99) by $\mathfrak{G}^{\theta_\nu}$ (Kronecker product). We can express both sides as sums of representations (98) again when we use repeatedly the relations

$$\mathfrak{G}^0 \times \mathfrak{G}^{\theta_\nu} = \mathfrak{G}^{\theta_\nu}, \quad \mathfrak{G}^{(r)\theta_\nu} \times \mathfrak{G}^{\theta_\nu} \leftrightarrow \mathfrak{G}^{(r-1)\theta_\nu} + \mathfrak{G}^{(r+1)\theta_\nu}, \qquad (r = 1, 2, \cdots, p-1)$$

$$\mathfrak{G}^{(p-1)\theta_\nu} \times \mathfrak{G}^{\theta_\nu} \leftrightarrow \mathfrak{G}^{\theta_{\nu+1}} + 2\mathfrak{G}^{(p-2)\theta_\nu}.$$

After the multiplication, $H(r_0, \cdots, r_{i-1}, r_i - 1, r_{i+1}, \cdots, r_{a-1})$ will appear on the left side of (99), whereas this term cannot appear on the right side. Because of the uniqueness of the irreducible constituents, we obtain a contradiction. This shows that the representations (98) are all the modular irreducible representations of $\mathfrak{G}$.

In the case of $GLH(2, p^a)$, we have to add a factor Δ^s, $0 \leqq s \leqq q - 2$, $\Delta = m_{11}m_{22} - m_{12}^2$, on the right side of (98), in order to obtain all the irreducible modular representations.

On the other hand, if (98) is to give a representation of the factor group $LF(2, p^a)$ of $SLH(2, p^a)$ modulo its centrum, then $-I$ must be represented by I in (98), i.e. the number $r_0 + r_1 + r_2 + \cdots + r_{a-1}$ must be even.[37]

31. The Cartan invariants and decomposition numbers (mod p) **of** $LF(2, p)$. We restrict ourselves to the case $LF(2, p)$, p an odd prime. The irreducible modular characters are here $\mathfrak{G}^{(0)}$, $\mathfrak{G}^{(2)}$, $\cdots$, $\mathfrak{G}^{(p-1)}$, the degrees are $1, 3, \cdots, p$. This shows, in particular, that the degree of the irreducible modular representations need not be a divisor of the order of the group. For the order of the radical we obtain

$$g - 1^2 - 3^2 - \cdots - p^2 = \frac{p(p^2-1)}{2} - \frac{(p+2)(p+1)p}{6} = \frac{p(p+1)(2p-5)}{6}.$$

[37] The modular characters of $GLH(3, p)$, $SLH(3, p)$ and $LF(3, p)$ have been determined by C. Mark in his Toronto thesis (to appear in the University of Toronto Studies).

There is no difficulty in computing the modular characters, and if they are arranged in the order $\mathfrak{G}^{(0)}, \mathfrak{G}^{(p-3)}, \mathfrak{G}^{(2)}, \mathfrak{G}^{(p-5)}, \cdots, \mathfrak{G}^{(p-3)/2}$ or $\mathfrak{G}^{(p-1)/2}, \mathfrak{G}^{(p-1)}$ we have

$$(100) \qquad C_1 = \begin{pmatrix} 2 & 1 & & & & & \\ 1 & 2 & 1 & & & & \\ & 1 & 2 & \cdot & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & 2 & 1 \\ & & & & & 1 & 3 \end{pmatrix} \qquad C_2 = (1)$$ [38]

for the C parts corresponding to the two blocks. (The coefficients not filled in are 0, there is just one 3 in the main diagonal of C_1). The first and the last u_κ both have the value p, all the other u_κ have the value $2p$.

Using formula (100) we can find D without any ambiguity. There must be two 1's in the first column. Beside them, we must have a 0 and a 1 in the second column etc. We thus find

$$D_1 = \begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ & 1 & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & \cdot \\ & & & 1 & 1 \\ & & & & 1 \\ & & & & 1 \end{pmatrix} \qquad D_2 = (1)$$

In this manner, we may obtain the values of the ordinary characters of $LF(2, p)$ except for the two p-singular classes. There is no difficulty in obtaining these missing values.[39]

UNIVERSITY OF TORONTO,
UNIVERSITY OF MICHIGAN.

---

[38] The determination of C according to this method is rather complicated, and we do not give the details of the computation. Using the methods sketched in op. cit. in footnote 20, C and D can be determined easily.

[39] These characters were first given by G. Frobenius, Sitzungsber. Preuss. Akad. 1896, p. 1013; the ordinary characters of the binary groups in $GF(p^a)$, $a > 1$ were first given by I. Schur, Jour. reine angew. Math. 132, 85, 1907, and independently by H. E. Jordan, Am. Jour. of Math. 29, 1907, p. 387.

Vol. 42, No. 3, July, 1941
ANNALS OF MATHEMATICS

ON POISSON'S SUMMATION FORMULA

By A. P. Guinand

(Received October 6, 1939)

1. Introduction

A number of papers have been written in recent years on various examples and applications of the formula

$$\tag{1.1} \tfrac{1}{2}f(0) + \sum_{n=1}^{\infty} f(n) = \int_0^{\infty} f(t)\,dt + 2\sum_{n=1}^{\infty}\int_0^{\infty} \cos 2\pi nt f(t)\,dt.$$

This formula is usually known as Poisson's summation formula. Other forms of the formula are

$$\tag{1.2} \sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty} \cos 2\pi nt f(t)\,dt,$$

and

$$\tag{1.3} \sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty} e^{2\pi int} f(t)\,dt.$$

Further, if we write

$$\tag{1.4} g(x) = 2\int_0^{\infty} \cos 2\pi xt f(t)\,dt,$$

then, with suitable conditions,

$$\tag{1.5} f(x) = 2\int_0^{\infty} \cos 2\pi xt g(t)\,dt$$

by Fourier's integral theorem, and (1.1) can be written

$$\tag{1.6} \tfrac{1}{2}f(0) + \sum_{n=1}^{\infty} f(n) = \tfrac{1}{2}g(0) + \sum_{n=1}^{\infty} g(n).$$

Several writers have pointed out that it is often difficult to justify the application of the formula even in cases where the formal result can be shown by other methods to be true.[1] Most methods which have been used for proving such formulae are elaborations of one of the two following types of formal procedure.

[1] Erdelyi (1), 408, Kober (1), 612.

(A). We may regard (1.2) as a Fourier series.[2] If we expand the periodic function

$$\sum_{n=-\infty}^{\infty} f(n + x)$$

as a Fourier series for $-\frac{1}{2} < x < \frac{1}{2}$, and then put $x = 0$ we formally obtain (1.2).
(B). We may regard (1.3) as a result in the theory of analytic functions.[3] If $f(z)$ is analytic, and we integrate $\pi \cot \pi z f(z)$ around the rectangle R with vertices $N + \frac{1}{2} \pm ai$, $-N - \frac{1}{2} \pm ai$, then by Cauchy's theorem

$$\frac{1}{2\pi i} \int_R \pi \cot \pi z f(z)\, dz = \sum_{n=-N}^{N} f(n).$$

Putting

$$\pi \cot \pi z = \pi i + 2\pi i \sum_{n=1}^{\infty} e^{-2n\pi i z}$$

when

$$I(z) < 0,$$

and

$$\pi \cot \pi z = -\pi i - 2\pi i \sum_{n=1}^{\infty} e^{2n\pi i z}$$

when

$$I(z) > 0,$$

and integrating term by term, we formally deduce (1.3) on making N tend to infinity and a to zero.

In addition to these methods two further methods have recently been developed which admit of application to functions having various types of singularities.
(C). Ferrar[4] derives a generalization of (1.1) by the use of Mellin transforms and the functional equation of the Riemann zeta-function. Conditions are applied to the Mellin transform of $f(x)$ rather than to $f(x)$ itself.
(D). In a recent paper[5] I gave another method for application to more general summation formulae of which (1.1) is a particular case. In this method we regard (1.1) as a Parseval theorem in the theory of Hankel transforms of functions of the Lebesgue class $L^2(0, \infty)$, and prove:

If $f(x)$ is an integral, and both $f(x)$ and $xf'(x)$ belong to $L^2(0, \infty)$, then

$$(1.7) \qquad \lim_{N\to\infty} \left\{ \sum_{n=1}^{N} f(n) - \int_0^N f(t)\, dt \right\} = \lim_{N\to\infty} \left\{ \sum_{n=1}^{N} g(n) - \int_0^N g(t)\, dt \right\},$$

[2] Courant and Hilbert (1), 64–65.
[3] Kronecker (1), Landsberg (1).
[4] Ferrar (1).
[5] Guinand (1). Put $a_n = 1$ in Theorem 2.

where $g(x)$ *satisfies*

$$\int_0^x g(y)\,dy = \frac{1}{\pi}\int_0^\infty \frac{\sin 2\pi xy}{y} f(y)\,dy$$

and is chosen so that it is the integral of its derivative.

The formula (1.7) obviously agrees with (1.6) if the series and integrals concerned converge. However, this is not necessary for the truth of (1.7), and it covers a number of examples in which (1.6) is inapplicable.

In the present paper we simplify the above result slightly, and extend it to functions of the Lebesgue classes $L^p(0, \infty)$, $(1 < p \leqq 2)$. We also give examples to show the ease with which the result can be applied, and discuss extensions to sums involving primitive characters.

2. Preliminary Lemmas

Lemma 1.[6] *If* $f(x)$ *belongs to* $L^p(0, \infty)$ *for some* p *satisfying* $1 < p \leqq 2$, *and*

$$p' = p/(p-1), \qquad R(\nu) > -1,$$

then there exists a function $g(x)$ *belonging to* $L^{p'}(0, \infty)$ *defined almost everywhere by*

$$g(x) = \operatorname*{l.i.m.}_{T\to\infty} 2\pi \int_0^T (xt)^{\frac{1}{2}} J_\nu(2\pi xt) f(t)\,dt. \tag{2.1}$$

Further

$$f(x) = \operatorname*{l.i.m.}_{T\to\infty} 2\pi \int_0^T (xt)^{\frac{1}{2}} J_\nu(2\pi xt) g(t)\,dt, \tag{2.2}$$

and if $f_1(x)$, $g_1(x)$ *and* $f_2(x)$, $g_2(x)$ *are two pairs of such transforms for the same values of* p *and* ν, *then*

$$\int_0^\infty f_1(x) g_2(x)\,dx = \int_0^\infty f_2(x) g_1(x)\,dx. \tag{2.3}$$

Lemma 2. *If* $f(x)$ *is an integral and tends to zero as* x *tends to infinity, and* $xf'(x)$ *belongs to* $L^p(0, \infty)$, $(p > 1)$ *then* $f(x)$ *belongs to* $L^p(0, \infty)$, *and* $x^{1/p}f(x)$ *tends to zero as* x *tends to* $+0$ *or infinity.*

Hardy[7] has proved that, if $\psi(x)$ belongs to $L^p(0, \infty)$, so does

$$\varphi(x) = \int_x^\infty \psi(t)\,\frac{dt}{t},$$

and

$$x^{1/p}\varphi(x) \to 0 \text{ as } x \to +0 \text{ or } x \to \infty.$$

Putting

$$\psi(x) = xf'(x)$$

the result follows immediately.

---

[6] Busbridge (1), or Kober (2). We only require the result for $\nu = \frac{1}{2}$.

[7] Titchmarsh (1), 396.

Lemma 3.[8] *The function* $x^{-1}\{[x] - x\}$ *is self-reciprocal with respect to the kernel* $2\pi x^{\frac{1}{2}} J_{\frac{1}{2}}(2\pi x)$. *That is*

$$2\pi \int_0^{\to\infty} \{[t] - t\}\, x^{\frac{1}{2}} t^{-\frac{1}{2}} J_{\frac{1}{2}}(2\pi xt)\, dt = x^{-1}\{[x] - x\}.$$

Lemma 4. *If* $f(x)$ *satisfies the conditions of Lemma* 2 *then it has a transform* $g(x)$ *with respect to the kernel* $2 \cos 2\pi x$ *given for* $x > 0$ *by*

$$g(x) = 2 \int_0^{\to\infty} \cos 2\pi xt f(t)\, dt. \tag{2.4}$$

Further

$$f(x) = 2 \int_0^{\to\infty} \cos 2\pi xt g(t)\, dt,$$

$g(x)$ *is an integral, both* $g(x)$ *and* $xg'(x)$ *belong to* $L^{p'}(0, \infty)$, $x^{1/p'}g(x)$ *tends to zero as* x *tends to* $+0$ *or* ∞, *and* $xf'(x)$, $xg'(x)$ *are a pair of transforms with respect to the kernel* $2\pi x^{\frac{1}{2}} J_{\frac{1}{2}}(2\pi x)$ *in the sense of Lemma* 1.

Put

$$G_1(x) = \underset{T\to\infty}{\text{l.i.m.}}\; 2 \int_0^{T} \cos 2\pi xt f(t)\, dt$$

and

$$G_2(x) = \underset{T\to\infty}{\text{l.i.m.}}\; 2\pi \int_0^{T} (xt)^{\frac{1}{2}} J_{\frac{1}{2}}(2\pi xt) t f'(t)\, dt.$$

By Lemmas 1 and 2 it follows that these expressions exist almost everywhere, and $G_1(x)$, $G_2(x)$ both belong to $L^{p'}(0, \infty)$. Now, if $x > 0$

$$\begin{aligned}\int_x^{\infty} | G_2(u) | \frac{du}{u} &\leqq \left\{\int_x^{\infty} | G_2(u) |^{p'} du\right\}^{1/p'} \left\{\int_x^{\infty} u^{-p}\, du\right\}^{1/p} \\ &\leqq \left\{\int_0^{\infty} | G_2(u) |^{p'} du\right\}^{1/p'} \left\{\frac{x^{1-p}}{p-1}\right\}^{1/p}.\end{aligned}$$

Hence the integral converges absolutely, and

$$\begin{aligned}\int_x^{\infty} G_2(u) \frac{du}{u} &= 2\pi \int_x^{\infty} \frac{du}{u} \underset{T\to\infty}{\text{l.i.m.}} \int_0^{T} (ut)^{\frac{1}{2}} J_{\frac{1}{2}}(2\pi ut) t f'(t)\, dt \\ &= 2\pi \lim_{T\to\infty} \int_x^{\infty} u^{-\frac{1}{2}}\, du \int_0^{T} t^{\frac{3}{2}} J_{\frac{1}{2}}(2\pi ut) f'(t)\, dt \\ &= 2\pi \lim_{T\to\infty} \int_0^{T} t^{\frac{3}{2}} f'(t)\, dt \int_x^{\infty} u^{-\frac{1}{2}} J_{\frac{1}{2}}(2\pi ut)\, du\end{aligned}$$

[8] Titchmarsh (2), 267. We also use the notation

$$\int_a^{\to\infty} = \lim_{T\to\infty} \int_a^{T}.$$

$$= \frac{1}{\pi x} \lim_{T\to\infty} \int_0^T \sin 2\pi xt f'(t)\, dt$$

$$= \frac{1}{\pi x} \lim_{T\to\infty} \left\{ [\sin 2\pi xt\, f(t)]_0^T - 2\pi x \int_0^T \cos 2\pi xt\, f(t)\, dt \right\}$$

$$= -2 \int_0^{\to\infty} \cos 2\pi xt\, f(t)\, dt.$$

The inversion of order of integration is justified by absolute convergence, and the integrated terms vanish by Lemma 2. Hence the limit (2.4) exists, and we may put

$$g(x) = -\int_x^\infty G_2(u)\, \frac{du}{u} = 2 \int_0^{\to\infty} \cos 2\pi xt\, f(t)\, dt$$

for all positive x. Hence $g(x)$ is an integral and $G_2(x) = xg'(x)$ almost everywhere. I.e. $xf'(x)$, $xg'(x)$ are transforms of the required type, and the rest of the lemma follows from Lemma 2.

3. The Summation Formula

We can now prove:

THEOREM 1. *If $f(x)$ is an integral, tends to zero at infinity, and $xf'(x)$ belongs to $L^p(0, \infty)$, $(1 < p \leqq 2)$ then*

$$\lim_{N\to\infty} \left\{ \sum_{n=1}^N f(n) - \int_0^N f(t)\, dt \right\} = \lim_{N\to\infty} \left\{ \sum_{n=1}^N g(n) - \int_0^N g(t)\, dt \right\} \tag{3.1}$$

where

$$g(x) = 2 \int_0^{\to\infty} \cos 2\pi x\, f(t)\, dt.$$

Now $x^{-1}\{[x] - x\}$ belongs to $L^p(0, \infty)$ for any $p > 1$, so by Lemmas 3 and 4 and (2.3)

$$\int_0^\infty x^{-1}\{[x] - x\}\, xf'(x)\, dx = \int_0^\infty x^{-1}\{[x] - x\}\, xf'(x)\, dx.$$

The left-hand side is

$$\int_0^\infty \{[x] - x\}\, f'(x)\, dx = (\{[x] - x\}\, f(x)]_0^\infty - \int_0^\infty f(x)\, d\{[x] - x\}$$

$$= -\lim_{N\to\infty} \int_0^N f(x)\, d\{[x] - x\}$$

$$= -\lim_{N\to\infty} \left\{ \sum_{n=1}^N f(n) - \int_0^N f(t)\, dt \right\}.$$

Treating the right-hand side in the same way, we have (3.1).

However, Theorem 1 does not cover functions having discontinuities, for $f(x)$, being an integral, is necessarily continuous. In order to obtain a sufficiently general theorem covering such functions we combine Theorem 1 with a known form of Poisson's formula.

THEOREM 2. *If $f(x)$ is of bounded variation in* $(0, \infty)$, *and tends to zero at infinity, then*

$$\lim_{N\to\infty}\left[\sum_{n=1}^{N}\tfrac{1}{2}\{f(n+0)+f(n-0)\}-\int_0^N f(t)\,dt\right]=\sum_{n=1}^{\infty}g(n)-\int_{\to 0}^{\to\infty}g(t)\,dt,$$

where

$$g(x)=2\int_{\to 0}^{\to\infty}\cos 2\pi xt\,f(t)\,dt$$

exists as a Cauchy integral at zero and infinity.

Titchmarsh[9] has shown that with these conditions

$$\lim_{N\to\infty}\left[\sum_{n=1}^{N}\tfrac{1}{2}\{f(n+0)+f(n-0)\}-\int_0^{N+\frac{1}{2}} f(t)\,dt\right]=\sum_{n=1}^{\infty}g(n)-\tfrac{1}{2}f(0+0).$$

Obviously

$$\lim_{N\to\infty}\int_N^{N+\frac{1}{2}} f(t)\,dt=0,$$

and we can show that[10]

$$\tfrac{1}{2}f(0+0)=\int_{\to 0}^{\to\infty}g(t)\,dt.$$

Theorem 2 follows immediately.

Now it sometimes happens that a function $f(x)$ satisfies the conditions of Theorem 1 for one set of values of p in the neighborhood of the origin and for another set of values at infinity. In order to cover all these possibilities we will take as our general theorem:

THEOREM 3. *If $f(x)$ can be expressed as the sum of three functions $f_1(x)$, $f_2(x)$, $f_3(x)$ such that all three tend to zero at infinity, $f_1(x)$ and $f_2(x)$ are integrals, $xf_1'(x)$ belongs to $L^{p_1}(0, \infty)$ for some p_1 in $(1 < p_1 \leqq 2)$, $xf_2'(x)$ belongs to $L^{p_2}(0, \infty)$ for some p_2 in $(1 < p_2 \leqq 2)$, and $f_3(x)$ is of bounded variation in $(0, \infty)$, then*

$$\lim_{N\to\infty}\left[\sum_{n=1}^{N}\tfrac{1}{2}\{f(n+0)+f(n-0)\}-\int_0^N f(t)\,dt\right]=\lim_{N\to\infty}\left[\sum_{n=1}^{N}g(n)-\int_{\to 0}^{N}g(t)\,dt\right], \tag{3.2}$$

[9] Titchmarsh (2), Theorem 45.

[10] Titchmarsh (2), Theorem 6. The required extension to the case $x = 0$ is trivial.

where

$$g(x) = 2\int_{\to 0}^{\to\infty} \cos 2\pi xt\, f(t)\, dt.$$

(3.2) holds for $f_1(x)$, $f_2(x)$ and $f_3(x)$ separately, and Theorem 3 follows by addition.

4. Examples

Many examples of Poisson's summation formula have been noted by various writers.[11] Most of these examples can be justified by Theorem 3 without difficulty. The examples given below are chosen to show the advantage of the use of Poisson's formula in the form (1.7) rather than in any of the forms (1.1), (1.2), or (1.3).

(i). Put

$$f(x) = x^{-s} \qquad (\tfrac{1}{2} < s < 1).$$

The conditions of Theorem 3 are satisfied if we put

$$f(x) = f_1(x) + f_2(x),$$

$$f_1(x) = \begin{cases} x^{-s} & (x < 1) \\ e^{1-x} & (x \geqq 1), \end{cases}$$

$$f_2(x) = \begin{cases} 0 & (x < 1) \\ x^{-s} - e^{1-x} & (x \geqq 1), \end{cases}$$

$$1 < p_1 < s^{-1}, \qquad p_2 = 2.$$

Hence

$$g(x) = 2\int_0^{\to\infty} t^{-s} \cos 2\pi xt\, dt = 2(2\pi)^{s-1} \sin \tfrac{1}{2}s\pi\Gamma(1-s)\, x^{s-1},$$

and

$$\lim_{N\to\infty}\left\{\sum_{n=1}^{N} n^{-s} - \frac{N^{1-s}}{1-s}\right\} = 2(2\pi)^{s-1} \sin \tfrac{1}{2}s\pi\Gamma(1-s) \lim_{N\to\infty}\left\{\sum_{n=1}^{N} n^{s-1} - \frac{N^{s}}{s}\right\}.$$

The expressions in brackets give the analytic continuation of $\zeta(s)$, and we have

$$\zeta(s) = 2(2\pi)^{s-1} \sin \tfrac{1}{2}s\pi\Gamma(1-s)\zeta(1-s),$$

the functional equation of the Riemann zeta-function.[12]

(ii). Consider

$$\varphi(z) = z^{-\frac{1}{2}} e^{\pi z/12} \sum_{n=0}^{\infty} p(n) e^{-2\pi n z} \qquad (z > 0),$$

[11] Cf. Erdelyi (1), Kober (1), Bochner (1), 37–38, and Titchmarsh (2), 64–65.

[12] Various methods of deriving this result from Poisson's formula have been given by other writers, but none is as direct as the above. Cf. Titchmarsh (2), 65.

where $p(n)$ is the number of unrestricted partitions of n. Using the generating function[13] of $p(n)$, we have

$$\varphi(z) = z^{-\frac{1}{4}} e^{\pi z/12} \prod_{n=1}^{\infty} (1 - e^{-2\pi n z})^{-1},$$

and

$$\log \varphi(z) = -\tfrac{1}{4} \log z + \frac{\pi z}{12} - \sum_{n=1}^{\infty} \log (1 - e^{-2\pi n z}).$$

Now put

$$f(x) = -\log (1 - e^{-2\pi x z})$$

in Theorem 1. The conditions are satisfied with $p = 2$, and

$$\begin{aligned} g(x) &= -2 \int_0^{\infty} \log (1 - e^{-2\pi t z}) \cos 2\pi x t \, dt \\ &= 2 \sum_{m=1}^{\infty} \frac{1}{m} \int_0^{\infty} e^{-2\pi m t z} \cos 2\pi x t \, dt \\ &= \frac{z}{\pi} \sum_{m=1}^{\infty} (m^2 z^2 + x^2)^{-1}. \end{aligned}$$

The term by term integration is justified by absolute convergence. If we apply Theorem 1 to this sum, we find that for $x > 0$

$$\begin{aligned} g(x) &= \frac{1}{x} \sum_{m=1}^{\infty} e^{-2\pi m x/z} + \frac{1}{2x} - \frac{z}{2\pi x^2} \\ &= \frac{1}{x} (e^{2\pi x/z} - 1)^{-1} + \frac{1}{2x} - \frac{z}{2\pi x^2}. \end{aligned}$$

Further

$$g(0) = \frac{z}{\pi} \sum_{m=1}^{\infty} \frac{1}{m^2 z^2} = \frac{\pi}{6z}.$$

Hence

$$\begin{aligned} -\sum_{n=1}^{\infty} \log (1 - e^{-2\pi n z}) - \frac{\pi}{12z} \\ = \lim_{N \to \infty} \Bigg[\sum_{n=1}^{\infty} \left\{ \frac{1}{n} (e^{2\pi n/z} - 1)^{-1} + \frac{1}{2n} - \frac{z}{2\pi n^2} \right\} \\ - \int_0^{N} \left\{ \frac{1}{t} (e^{2\pi t/z} - 1)^{-1} + \frac{1}{2t} - \frac{z}{2\pi t^2} \right\} dt \Bigg] \\ = \sum_{n=1}^{\infty} \frac{1}{n} (e^{2\pi n/z} - 1)^{-1} + \tfrac{1}{2}\gamma - \frac{\pi z}{12} \\ - \lim_{N \to \infty} \left[\int_0^{N} \left\{ \frac{1}{t} (e^{2\pi t/z} - 1)^{-1} + \frac{1}{2t} - \frac{z}{2\pi t^2} \right\} dt - \tfrac{1}{2} \log N \right]. \end{aligned}$$

[13] Hardy and Wright (1), 272.

Now

$$\sum_{n=1}^{\infty}\frac{1}{n}(e^{2\pi n/z}-1)^{-1}=\sum_{n=1}^{\infty}\frac{1}{n}\sum_{m=1}^{\infty}e^{-2\pi mn/z}$$
$$=\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{1}{n}e^{-2\pi mn/z}$$
$$=-\sum_{m=1}^{\infty}\log(1-e^{-2\pi m/z}),$$

and the above limit is equal to

$$\int_0^z\left\{\frac{1}{t}(e^{2\pi t/z}-1)^{-1}+\frac{1}{2t}-\frac{z}{2\pi t^2}\right\}dt+\int_z^{\infty}\left\{\frac{1}{t}(e^{2\pi t/z}-1)^{-1}-\frac{z}{2\pi t^2}\right\}dt$$
$$+\frac{1}{2}\lim_{N\to\infty}\left[\int_z^N\frac{dt}{t}-\log N\right]$$
$$=\int_0^1\left\{\frac{1}{u}(e^{2\pi u}-1)^{-1}+\frac{1}{2u}-\frac{1}{2\pi u^2}\right\}du$$
$$+\int_1^{\infty}\left\{\frac{1}{u}(e^{2\pi u}-1)^{-1}-\frac{1}{2\pi u^2}\right\}du-\tfrac{1}{2}\log z$$
$$=K-\tfrac{1}{2}\log z,$$

where K is a constant. Hence

$$-\sum_{n=1}^{\infty}\log(1-e^{-2\pi nz})-\frac{\pi}{12z}=-\sum_{m=1}^{\infty}\log(1-e^{-2\pi m/z})-\frac{\pi z}{12}+\tfrac{1}{2}\gamma-K+\tfrac{1}{2}\log z.$$

That is

$$\log\varphi(z)=\log\varphi\left(\frac{1}{z}\right)+\tfrac{1}{2}\gamma-K.$$

Putting $z=1$ it follows that $K=\frac{1}{2}\gamma$, and

$$\varphi(z)=\varphi\left(\frac{1}{z}\right).$$

Hence

$$z^{-\frac{1}{4}}e^{\pi z/12}\sum_{n=0}^{\infty}p(n)e^{-2\pi nz}=z^{\frac{1}{4}}e^{\pi/12z}\sum_{n=0}^{\infty}p(n)e^{-2\pi n/z}.$$

This is a well-known result in the theories of partitions and elliptic modular functions.

(iii). Put[14]

$$f(x)=(z^2+x^2)^{-\frac{1}{2}}$$

[14] Cf. Watson (1).

then the conditions of Theorem 1 are satisfied with $p = 2$, and

$$g(x) = 2\int_0^{\to\infty} (z^2 + t^2)^{-\frac{1}{2}} \cos 2\pi xt\, dt = 2K_0(2\pi xz).$$

Hence

$$\begin{aligned} 2\sum_{n=1}^{\infty} K_0(2\pi nz) - \frac{1}{2z} &= \lim_{N\to\infty}\left[\sum_{n=1}^{N} (z^2 + n^2)^{-\frac{1}{2}} - \int_0^{N} (z^2 + t^2)^{-\frac{1}{2}}\, dt\right] \\ &= \lim_{N\to\infty}\left[\sum_{n=1}^{N}\left\{(z^2 + n^2)^{-\frac{1}{2}} - \frac{1}{n}\right\} + \left\{\sum_{n=1}^{N}\frac{1}{n} - \log N\right\}\right. \\ &\qquad\qquad \left. + \left\{\log N - \int_0^{N/z} (1 + u^2)^{-\frac{1}{2}}\, du\right\}\right] \\ &= \sum_{n=1}^{\infty}\left\{(z^2 + n^2)^{-\frac{1}{2}} - \frac{1}{n}\right\} + \gamma - \lim_{N\to\infty}\left[\log\left\{\frac{1}{z} + \left(\frac{1}{z^2} + \frac{1}{N^2}\right)^{\frac{1}{2}}\right\}\right] \\ &= \sum_{n=1}^{\infty}\left\{(z^2 + n^2)^{-\frac{1}{2}} - \frac{1}{n}\right\} + \gamma + \log \tfrac{1}{2}z. \end{aligned}$$

5. Series Involving Primitive Characters

Poisson's summation formula can be generalized by the introduction of the primitive characters as coefficients.[15] We require the following elementary properties of characters:[16]

(i). *If $\chi(n)$ is a character modulo k ($k > 1$) then*

$$\chi(m)\chi(n) = \chi(mn),$$

$$\chi(m + rk) = \chi(m)$$

where k, m, n, and r are integers. Further $\chi(n)$ vanishes if n is not prime to k.

(ii). *If $\chi(n)$ is a primitive character modulo k ($k > 1$) then*

$$\sum_{n=1}^{k} \chi(n) e^{2\pi mni/k} = \epsilon(\chi) k^{\frac{1}{2}} \bar{\chi}(m), \tag{5.1}$$

where

$$\epsilon(\chi) = k^{-\frac{1}{2}} \sum_{n=1}^{k} \chi(n) e^{2\pi ni/k}, \tag{5.2}$$

and $\bar{\chi}(m)$ is the conjugate of $\chi(m)$.

Now $\chi(-1) = \pm 1$, since

$$\{\chi(-1)\}^2 = \chi(1) = 1.$$

---

[15] Some particular cases of this type of generalization have been given by Ramanujan. Cf. Titchmarsh (2), 82.

[16] Landau (1), chapters XXII and XXX. The notation for $\epsilon(\chi)$ used here differs from Landau's by a factor $-i$ in the case $\chi(-1) = -1$.

Suppose that $\chi(-1) = 1$; then

$$\sum_{n=1}^{k} \chi(n) \sin \frac{2\pi mn}{k} = 0,$$

since

$$\chi(k - r) = \chi(-r) = \chi(-1)\chi(r) = \chi(r),$$

and consequently the terms $n = r$ and $n = k - r$ cancel. Also the term $n = \frac{1}{2}k$, if it occurs, is zero since $\chi(\frac{1}{2}k) = 0$. Hence (5.1) becomes

$$\sum_{n=1}^{k} \chi(n) \cos \frac{2\pi mn}{k} = \epsilon(\chi)k^{\frac{1}{2}}\bar{\chi}(m). \tag{5.3}$$

Now if we put

$$f(x) = \begin{cases} \cos \dfrac{2\pi mx}{k} & (-y < x < y), \\ 0 & (|x| > y) \end{cases}$$

in (1.2), we find that

$$\sum_{-y \leq n \leq y}{}' \cos \frac{2\pi mn}{k} = \frac{k}{\pi} \sum_{n=-\infty}^{\infty} \frac{1}{nk + m} \sin \frac{2\pi y}{k}(nk + m),$$

where the prime indicates that the terms $n = \pm y$ are to be halved if y is an integer. If we multiply this result by $\chi(m)$, and sum from $m = 1$ to $m = k$, we have

$$\sum_{m=1}^{k} \chi(m) \sum_{-y \leq n \leq y}{}' \cos \frac{2\pi mn}{k} = \frac{k}{\pi} \sum_{m=1}^{k} \chi(m) \sum_{n=-\infty}^{\infty} \frac{1}{nk + m} \sin \frac{2\pi y}{k}(nk + m).$$

By (5.3) this is

$$\begin{aligned} \epsilon(\chi)k^{\frac{1}{2}} \sum_{-y \leq n \leq y}{}' \bar{\chi}(n) &= \frac{k}{\pi} \sum_{m=1}^{k} \sum_{n=-\infty}^{\infty} \frac{\chi(nk + m)}{nk + m} \sin \frac{2\pi y}{k}(nk + m) \\ &= \frac{k}{\pi} \sum_{n=-\infty}^{\infty} \frac{\chi(n)}{n} \sin \frac{2\pi ny}{k}. \end{aligned}$$

Hence

$$\epsilon(\chi)k^{\frac{1}{2}} \sum_{1 \leq n \leq y}{}' \bar{\chi}(n) = \frac{k}{\pi} \sum_{n=1}^{\infty} \frac{\chi(n)}{n} \sin \frac{2\pi ny}{k}.$$

Now let us consider the integral

$$2\pi k^{-\frac{1}{2}} \int_0^{\to\infty} t^{-1} \Big\{ \sum_{1 \leq n \leq t}{}' \chi(n) \Big\} \left(\frac{xt}{k}\right)^{\frac{1}{2}} J_{\frac{1}{2}}\left(\frac{2\pi xt}{k}\right) dt. \tag{5.4}$$

We have

$$\frac{2\pi x^{\frac{1}{2}}}{k}\int_0^N \Big\{\sum_{1\leq n\leq t}{}' \chi(n)\Big\} t^{-\frac{1}{2}} J_{\frac{1}{2}}\left(\frac{2\pi x t}{k}\right) dt$$

$$= \frac{2\pi x^{\frac{1}{2}}}{k}\sum_{n=1}^{N-1}\{\chi(1)+\chi(2)+\cdots+\chi(n)\}\int_n^{n+1} t^{-\frac{1}{2}} J_{\frac{1}{2}}\left(\frac{2\pi x t}{k}\right) dt$$

$$= \frac{k^{\frac{1}{2}}}{\pi x}\sum_{n=1}^{N-1}\{\chi(1)+\chi(2)+\cdots+\chi(n)\}\left\{\frac{1}{n}\sin\frac{2\pi n x}{k} - \frac{1}{n+1}\sin\frac{2\pi(n+1)x}{k}\right\}$$

$$= \frac{k^{\frac{1}{2}}}{\pi x}\sum_{n=1}^{N}\frac{\chi(n)}{n}\sin\frac{2\pi n x}{k} - \frac{k}{\pi N x}\sin\frac{2\pi N x}{k}\sum_{n=1}^{N}\chi(n).$$

If we make N tend to infinity the integral (5.4) becomes

$$\frac{k^{\frac{1}{2}}}{\pi x}\sum_{n=1}^{\infty}\frac{\chi(n)}{n}\sin\frac{2\pi n x}{k} = \epsilon(\chi)x^{-1}\Big\{\sum_{1\leq n\leq x}{}' \bar{\chi}(n)\Big\}.$$

That is the functions

$$x^{-1}\Big\{\sum_{1\leq n\leq x}{}' \chi(n)\Big\}, \qquad \epsilon(\chi)x^{-1}\Big\{\sum_{1\leq n\leq x}{}' \bar{\chi}(n)\Big\}$$

are transforms with respect to the kernel

$$\frac{2\pi x^{\frac{1}{2}}}{k} J_{\frac{1}{2}}\left(\frac{2\pi x}{k}\right).$$

If we now apply the methods of section 3 we can obtain a result analogous to Theorem 1. We can also obtain a result corresponding to the case $\chi(-1) = -1$. We have:

THEOREM 4. *If $\chi(n)$ is a primitive character modulo k ($k > 1$), $f(x)$ is an integral, tends to zero at infinity, $xf'(x)$ belongs to $L^p(0, \infty)$, ($1 < p \leq 2$), and either*
(i) $\chi(-1) = 1$,

$$g(x) = 2\epsilon(\chi)k^{-\frac{1}{2}}\int_0^{\rightarrow\infty}\cos\frac{2\pi x t}{k} f(t)\,dt,$$

or (ii) $\chi(-1) = -1$,

$$g(x) = -2i\epsilon(\chi)k^{-\frac{1}{2}}\int_0^{\rightarrow\infty}\sin\frac{2\pi x t}{k} f(t)\,dt,$$

then

$$\sum_{n=1}^{\infty}\chi(n)f(n) = \sum_{n=1}^{\infty}\bar{\chi}(n)g(n).$$

For real primitive characters $\bar{\chi}(n) = \chi(n)$, and it follows on substituting $f(x) = e^{-\pi x^2/k} = g(x)$ in case (i) that $\epsilon(\chi) = 1$. Similarly in case (ii) $f(x) = xe^{-\pi x^2/k} = g(x)$ gives $-i\epsilon(\chi) = 1$, and we have:

THEOREM 5. *If, in addition to the assumptions of Theorem* 4, $\chi(n)$ *is a real character, and either*
(i) $\chi(-1) = 1$,

$$g(x) = 2k^{-\frac{1}{2}} \int_0^{\to\infty} \cos \frac{2\pi xt}{k} f(t)\, dt,$$

or (ii) $\chi(-1) = -1$,

$$g(x) = 2k^{-\frac{1}{2}} \int_0^{\to\infty} \sin \frac{2\pi xt}{k} f(t)\, dt,$$

then

$$\sum_{n=1}^{\infty} \chi(n)f(n) = \sum_{n=1}^{\infty} \chi(n)g(n).$$

PRINCETON, NEW JERSEY.

BIBLIOGRAPHY

S. Bochner (1), Vorlesungen über Fouriersche Integrale, (Leipzig 1932).

I. W. Busbridge (1), A theory of general transforms for functions of the class $L^p(0, \infty)$ $(1 < p \leqq 2)$, Quart. J. of Math. (Oxford), 9 (1938), 148–160.

R. Courant and D. Hilbert (1), Methoden der mathematischen Physik I, (Berlin 1931).

A. Erdelyi (1), Gewisse Reihentransformationen, die mit der linearen Transformationsformel der Theta-funktion zusammenhangen, Compositio Math. 4 (1937), 406–423.

W. L. Ferrar (1), Summation formulae and their relation to Dirichlet's series II, Compositio Math. 4 (1937), 394–405.

A. P. Guinand (1), Summation formulae and self-reciprocal functions, Quart. J. of Math. (Oxford), 9 (1938), 53–67.

G. H. Hardy and E. M. Wright (1), Introduction to the theory of numbers, (Oxford 1938).

H. Kober (1), Transformationsformeln gewisser Besselscher Reihen. Beziehungen zu Zeta-funktionen, Math. Zeitschrift 39 (1934), 609–624. (2), Hankelsche Transformationen, Quart. J. of Math. (Oxford), 8 (1937), 186–199.

L. Kronecker (1), Bemerkungen über die Darstellung von Reihen durch Integrale, Journal für Math. 105 (1889), 157–159 and 345–354.

E. Landau (1), Handbuch der Lehre von der Verteilung der Primzahlen, (Leipzig 1909).

G. Landsberg (1), Zur Theorie Gausschen Summen und der linearen Transformation der Theta-funktionen, Journal für Math. 111 (1893), 234–253.

E. C. Titchmarsh (1), Theory of functions, (Oxford 1939). (2), Introduction to the theory of Fourier integrals, (Oxford 1937).

G. N. Watson (1), Some self-reciprocal functions, Quart. J. of Math. (Oxford), 2 (1931), 298–309.

Annals of Mathematics
Vol. 42, No. 3, July, 1941

ON THE REARRANGEMENT OF CONDITIONALLY CONVERGENT SERIES

By C. T. Rajagopal

(Received June 25, 1940)

This note is concerned with an application of the idea which finds expression in

Littlewood's Theorem [2].[1] *If $F(x)$ is positive monotone decreasing and $\{D_n\}$ is a sequence such that*

$$\left.\begin{aligned} 0 = D_0 < D_1 < D_2 < \cdots < D_n \to \infty, \\ D_n - D_{n-1} = O(1), \end{aligned}\right\} \quad n \to \infty,$$

then $\sum^\infty (D_n - D_{n-1})F(D_n)$ converges or diverges with $\int^\infty F(x)\,dx$.

In an earlier note [6], I have shown that the idea behind the theorem—that of a condensation effected on series by integrating a suitable function over intervals of the type (D_{n-1}, D_n)—supplies the basis for a unified theory of convergence criteria for series of positive terms, more comprehensive than Pringsheim's [3] in certain respects. In another note [7], I have shown that the idea suggests elegant proofs of certain classic theorems of Pringsheim [4]. I prove here that the same idea can be used to generalize the enunciations and simplify the proofs of Pringsheim's theorems concerning the rearrangement of conditionally convergent series [5].[2]

1. Notation

$S \equiv \sum_{n=1}^\infty (-1)^{n-1}a_n (a_n > 0)$ is a conditionally convergent series, $s_n = \sum_{\nu=1}^{n} (-1)^{\nu-1}a_\nu$. $\sum_{n=1}^\infty (-1)^{n-1}a_n'$ is a rearrangement of s in which $s'_{p+q} = \sum_{\nu=1}^{p+q} (-1)^{\nu-1}a'_\nu$ contains the first p positive terms and the first q negative terms of s.

1.1. The theorems which follow are typical of the results obtained by Pringsheim. It is supposed in their enunciations that $a_n \equiv a(n)$, where $a(x)$ is a differentiable function of x, and that limiting relations exist, as specified, between p and q.

Theorem A. *If, in $S \equiv \sum_1^\infty (-1)^{n-1}a_n$, a_n^{-1} and na_n tend steadily to infinity with n, then $\sum_1^\infty (-1)^{n-1}a_n'$ has a sum s' given by*

$$s' = s + \lim_{p > q \to \infty} (p - q)a_{2q}.$$

[1] The numbers in [ ] denote the references at the end of this note. A slight generalization of Littlewood's theorem, which dispenses with the condition: $D_n - D_{n-1} = O(1)$, is given in de la Vallée-Poussin, *Cours d'analyse*, 1 (Louvain-Paris, 1926), 398. A generalization in a different direction is given in Rajagopal [8], §2.

[2] All the references to Pringsheim in the sequel are to [5].

THEOREM B. *If, in* s, $\lim_{n\to\infty} na_n$ *is finite, equal to* g *(say), then*

$$s' = s + \frac{g}{2} \lim_{p,q\to\infty} (\log 2p - \log 2q).$$

THEOREM B′. *If, in* s, $\lim_{n\to\infty} n\cdot l_1 n\cdot l_2 n \cdots l_{x-1}n\cdot a_n = g$ *(where* $l_1 n = \log n$, $l_2 n = \log\log n, \cdots$*), then*

$$s' = s + \frac{g}{2} \lim_{p,q\to\infty} (l_x 2p - l_x 2q).$$

THEOREM C. *If, in* s, na_1 *tends steadily to the limit zero and if* $(p - q)a_{2q} = O(1)$ $(p > q \to \infty)$, *then* $s' = s$.

Theorems A, B, C, B′ are suggested by (26), (30), (31), (34) respectively in Pringsheim's paper. The restriction on a_n in all the theorems is of the form: $\lim_{n\to\infty} f(n)\cdot a_n$ is finite, where $f(x)$ is a positive differentiable function for all large x. The distinction between the theorems arises from the fact that in Theorem A, $\lim f(n)/n = 0$; in Theorem B, $\lim f(n)/n$ is finite and non-zero; in Theorems B′ and C, $\lim f(n)/n = \infty$. A question, suggested by Littlewood's theorem, is whether we cannot with advantage replace $f(n)$ by $f(D_n)$ in the condition imposed on a_n. An attempt at an answer is contained in Theorems 1 and 1a which are found to cover all the cases discussed by Pringsheim.

In all the theorems and the lemmas proved below it is tacitly assumed that $0 < d_n \equiv D_n - D_{n-1} < K$ *(fixed).*

2

THEOREM 1. *If* $f(x)$ *is a positive function with a continuous derivative* $f'(x)$ *such that* $\dfrac{f'(x)}{f(x)} = o(1)$ $(x \to \infty)$, *and if* $\lim_{n\to\infty} \dfrac{a_n\cdot f(D_n)}{D_n - D_{n-2}} = g$, *then* $\sum_1^\infty (-1)^{n-1}a'_n$ *oscillates between* $s + g \overline{\underline{\lim}}_{p,q\to\infty} \int_{D_{2q}}^{D_{2p}} \dfrac{dx}{f(x)}$, *provided* $\int_{D_{2q}}^{D_{2p}} \dfrac{dx}{f(x)}$ *is bounded.*

PROOF:

$$\frac{D_\nu - D_{\nu-2}}{f(D_\nu)} = \int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} + \int_{D_{\nu-2}}^{D_\nu} \left[\frac{1}{f(D_\nu)} - \frac{1}{f(x)}\right] d(x - D_{\nu-2}).$$

Integrating by parts the last term of this identity, we find

$$\begin{aligned}\frac{D_\nu - D_{\nu-2}}{f(D_\nu)} &= \int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} + \int_{D_{\nu-2}}^{D_\nu} (x - D_{\nu-2})\, d\left[\frac{1}{f(x)}\right] \\ &= \int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} + o(1)\int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} \qquad (\nu \to \infty);\end{aligned}$$

whence, given $\epsilon > 0$ arbitrarily small, we can find n_0 such that

$$(1) \qquad (g - \epsilon) \int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} < a_\nu < (g + \epsilon)\int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} \qquad (\nu \geqq n_0).$$

If for the sake of definiteness, we suppose $p > q$,

(2) $$\sum_{\nu=1}^{p+q} (-1)^{\nu-1} a'_\nu = \sum_{\nu=1}^{2q} (-1)^{\nu-1} a_\nu + (a_{2q+1} + a_{2q+3} + \cdots + a_{2p-1}).$$

Hence, if $2q + 1 \geqq n_0$, summing (1) for $\nu = 2q + 1, 2q + 3, \cdots 2p - 1$, we obtain from (2),

(3) $$s'_{p+q} \begin{cases} < s_{2q} + (g + \epsilon) \displaystyle\int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)}, \\ > s_{2q} + (g - \epsilon) \displaystyle\int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)}, \end{cases}$$

where we can write

(4) $$\int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)} = \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} + \int_{D_{2q-1}}^{D_{2p}} \frac{dx}{f(x)} - \int_{D_{2p-1}}^{D_{2p}} \frac{dx}{f(x)}$$

and show that the second and third terms of the right hand member are $o(1)$ as $p, q \to \infty$.

First, when $0 < v - u = O(1)$ $(u \to \infty)$,

$$\log \frac{f(v)}{f(u)} = \int_u^v \frac{f'(x)}{f(x)} dx = o(1) \qquad (u \to \infty).$$

Or,

(5) $$f(v) \sim f(u).$$

Next,

$$\int_{D_{2n-1}}^{D_{2n}} \frac{dx}{f(x)} = \frac{D_{2n} - D_{2n-1}}{f\{D_{2n-1} + \theta(D_{2n} - D_{2n-1})\}} \qquad (0 < \theta < 1)$$

$$\sim \frac{D_{2n} - D_{2n-1}}{f(D_{2n})} \qquad (n \to \infty),$$

in virtue of (5).

Also,

$$0 < \frac{D_{2n} - D_{2n-1}}{f(D_{2n})} < \frac{D_{2n} - D_{2n-2}}{f(D_{2n})} = o(1) \qquad (n \to \infty),$$

in consequence of the restriction $\dfrac{D_N - D_{N-2}}{f(D_N)} = o(1)$ $(N \to \infty)$ which results from $a_N = o(1)$ and $\dfrac{a_N}{(D_N - D_{N-2})/f(D_N)} \sim g$; whence

(6) $$\int_{D_{2n-1}}^{D_{2n}} \frac{dx}{f(x)} = o(1) \qquad (n \to \infty).$$

We can now complete the proof by using (6) in (4) and (4) in (3).

In certain applications it is useful to have the following complement to Theorem 1.

THEOREM 1a. *If, in Theorem 1, we replace the restriction relating to $f'(x)$ by the assumption that $f(x)$ tends steadily to ∞ with x, then either $s'_{p+q} - s_{2q}$ $(p > q)$, or $s'_{p+q} - s_{2p}$ $(q > p)$, has an asymptotic representation which oscillates between $g \overline{\underline{\lim}}_{p,q\to\infty} \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)}$, provided $\int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)}$ is bounded.*

PROOF: Supposing (as before) $p > q$,

$$\frac{D_{2q+1} - D_{2q-1}}{f(D_{2q+1})} + \frac{D_{2q+3} - D_{2q+1}}{f(D_{2q+3})} + \cdots + \frac{D_{2p-1} - D_{2p-3}}{f(D_{2p-1})} = \int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)} + o(1)$$

$$(q \to \infty).$$

As in (4), we can split $\int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)}$ into three terms of which the last two are $o(1)$ as $p, q \to \infty$, in virtue of the condition: $f(x) \to \infty$ with x.

Hence, as $q \to \infty$,

$$a_{2q+1} + a_{2q+3} + \cdots + a_{2p-1} \sim g\left[\frac{D_{2q+1} - D_{2q-1}}{f(D_{2q+1})} + \cdots + \frac{D_{2p-1} - D_{2p-3}}{f(D_{2p-1})}\right]$$

$$= g \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} + o(1)$$

which oscillates between $g \overline{\underline{\lim}} \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)}$.

REMARK. The argument employed to prove Theorems 1 and 1a shows that, under the conditions of these theorems, the sum s' of the rearranged series is given by

$$s' = s + g \lim \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} \tag{7}$$

provided this limit exists and is finite. If the limit is infinite, then $\sum_1^\infty (-1)^{n-1} a'_n$ diverges.

2.1. We may consider, in the light of this remark, series of the type in Pringsheim's paper, in forms slightly more general than his. The generalized series are all of the kind whose conditional convergence we can establish by

LEMMA 1. *If, as $x \to \infty$, $f(x)$ decreases steadily to the limit zero and $\int^\infty F(x)\, dx$ is divergent, then $\sum^\infty (-1)^{n-1} F(D_n)$ is conditionally convergent.*

PROOF: The convergence of the series in question follows from Leibnitz's rule for alternating series.

To prove that the convergence is not absolute, we observe that $KF(D_n) > (D_n - D_{n-1})F(D_n)$ which, by Littlewood's theorem, is the general term of a divergent series.

Examples of the lemma are furnished by $\sum^{\infty} (-1)^{n-1} a_n$, where a_n may be any one of the functions:

$$\frac{1}{D_n^{\mu}} \; (0 < \mu \leqq 1), \quad \frac{1}{\log D_n}, \quad \frac{\log D_n}{D_n}, \quad \frac{1}{D_n \cdot \log D_n}, \quad \frac{1}{AD_n + BD_n^{\mu}} \quad \begin{pmatrix} A, B > 0; \\ 1 > \mu > 0 \end{pmatrix}.$$

A second proposition from which we can infer the conditional convergence of a certain type of alternating series is

LEMMA 2. *If* $u_n > 0$ *is such that*

$$\frac{u_n/d_n}{u_{n-1}/d_{n-1}} = 1 - \mu \frac{d_n}{D_n} - O\left(\frac{d_n}{D_n^{\lambda}}\right) \qquad (\lambda > 1), \tag{8}$$

then (i) $\sum^{\infty} u_n$ *is convergent or divergent according as* $\mu > 0$, *or* $\leqq 1$; (ii) $\sum^{\infty} (-1)^{n-1} u_n/d_n$ *is convergent for* $\mu > 0$, (iii) *the convergence in* (ii) *is conditional for* $0 < \mu \leqq 1$.

This result is at once an extension and an amplification of Gauss's test.

(i) and (ii) are immediate consequences of the relation: $\dfrac{u_n}{d_n} \sim c D_n^{-\mu}$, equivalent to (8)[3].

(iii) follows from the fact that $u_n/d_n > u_n/k$ which, by (i), is the general term of a divergent series when $\mu \leqq 1$.

As an application of Lemma 2, consider the case:

$$u_n = d_n \frac{\prod_{\nu=1}^{n} (\alpha d_\nu + D_\nu)(\beta d_\nu + D_\nu)}{\prod_{\nu=1}^{n} (\gamma d_\nu + D_\nu)(\delta d_\nu + D_\nu)}$$

where $\alpha, \beta, \gamma, \delta$ are real and the two latter $\neq -D_n/d_n$ $(n = 1, 2, \cdots)$. It is clear that $\sum u_n$ reduces to the hypergeometric type when $d_n = 1$, $\delta = 0$. Further,

$$\frac{u_n/d_n}{u_{n-1}/d_{n-1}} = 1 - (\gamma + \delta - \alpha - \beta) \frac{d_n}{D_n} - O\left(\frac{d_n}{D_n^2}\right).$$

The terms of $\sum^{\infty} u_n$ being ultimately of the same sign and the terms of $\sum^{\infty} (-1)^{n-1} u_n/d_n$ alternately positive and negative, Lemma 2 shows that

$$\sum^{\infty} (-1)^{n-1} a_n \equiv \sum^{\infty} (-1)^{n-1} \frac{\prod_{\nu=1}^{n} (\alpha d_\nu + D_\nu)(\beta d_\nu + D_\nu)}{\prod_{\nu=1}^{n} (\gamma d_\nu + D_\nu)(\delta d_\nu - D_\nu)} \tag{9}$$

is conditionally convergent for $0 < \gamma + \delta - \alpha - \beta \leqq 1$.

REMARK. a_n in (9) can be related to a generalization of Gauss's function:

$$\left.\begin{aligned} \Pi(z) &= \lim_{n \to \infty} \frac{n^z}{(1 + z)\left(1 + \frac{z}{2}\right) \cdots \left(1 + \frac{z}{n}\right)} \\ &= \Gamma(1 + z) \end{aligned}\right\} \quad (z \neq -1, -2, \cdots)$$

---

[3] Lemma 2 (i) is, in fact, a particular case of a theorem first proved by me in *Bull. Amer. Math. Soc.* 43 (1937), 411. But the proof of the theorem given there is not applicable to Lemma 2 (ii).

whose reciprocal is, in Weierstrass's form:

$$\frac{1}{\Pi(z)} = e^{Cz} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right) e^{-z/n},$$

C being Euler's constant.

The generalization of $\Pi(z)$ referred to is due to Cesàro[4]. It is given by

$$\Pi_{(d_n)}(z) = \lim_{n\to\infty} \frac{D_n^z}{\left(1 + z\frac{d_1}{D_1}\right)\left(1 + z\frac{d_2}{D_2}\right)\cdots\left(1 + z\frac{d_n}{D_n}\right)} \tag{10}$$

$$\left(z \neq \frac{-D_1}{d_1}, \frac{-D_2}{d_2}, \dots\right).$$

Its reciprocal can be expressed in the form:

$$\frac{1}{\Pi_{(d_n)}(z)} = e^{C(d_n)\cdot z} \prod_{n=1}^{\infty} \left(1 + z\frac{d_n}{D_n}\right) e^{-zd_n/D_n}$$

where $C_{(d_n)} = \lim_{n\to\infty} \left(\sum_{\nu=1}^{n} \frac{d_\nu}{D_\nu} - \log D_n\right)$ is a constant with some of the elementary properties and application of C.[5]

(10) *shows that* a_n *in* (9) *is subject to the condition*

$$\lim_{n\to\infty} a_n \cdot D_n^{\gamma+\delta-\alpha-\beta} = \frac{\Pi_{(d_n)}(\gamma)\cdot\Pi_{(d_n)}(\delta)}{\Pi_{(d_n)}(\alpha)\cdot\Pi_{(d_n)}(\beta)}. \tag{11}$$

This fact will be used later (§3.1).

2.2. Taking the series in §2.1, we can find out how they are affected by particular rearrangements, supposing that in all of them appearing as examples, $d_n \sim k > 0$ $(n \to \infty)$.

Examples. 1) Let $\sum_1^\infty (-1)^{n-1} a_n \equiv \sum_1^\infty (-1)^{n-1}/D_n \cdot \log D_n$. Taking $f(x) = x \cdot \log x$ in Theorem 1, we find $g = 1/2k$. Hence, by (7),

$$s' - s = \frac{1}{2k} \lim_{p,q\to\infty} \int_{D_{2q}}^{D_{2p}} \frac{dx}{x \cdot \log x} = \frac{1}{2k} \lim \log\left(\frac{\log D_{2p}}{\log D_{2q}}\right)$$

provided this limit exists.

In particular, if $0 < \underline{\overline{\lim}}\, (D_{2p}/D_{2q}) < \infty$, there is no alternation in the sum of the series, due to the rearrangement.

Next, if $D_{2p} \sim D_{2q}^\lambda$ $(\lambda > 1)$, the alteration is $(\log \lambda)/2k$.

Finally, if $D_{2p} \sim \lambda^{D_{2q}}$ $(\lambda > 1)$, the rearranged series diverges to ∞.

2) Let $\sum_1^\infty (-1)^{n-1} a_n \equiv \sum_1^\infty (-1)^{n-1}/AD_n + BD_n^\mu$ $(A, B > 0; 1 > \mu > 0)$.

---

[4] The function is mentioned in E. T. Whittaker, *Modern Analysis* (Cambridge, 1902), 195. The notation in Whittaker is, however, different from the one employed here.

[5] E.g., Rajagopal [8], §3.

With $f(x) = Ax + Bx^\mu$, $g = 1/2k$ so that

$$s' - s = \frac{1}{2k} \lim_{p,q\to\infty} \int_{D_{2q}}^{D_{2p}} \frac{dx}{Ax + Bx^\mu} = \frac{1}{2kA(1-\mu)} \lim \log \frac{AD_{2p}^{1-\mu} + B}{AD_{2q}^{1-\mu} + B}$$

if this limit exists.

Thus, if $D_{2p} \sim bD_{2q}$, the alteration in the sum of the series is $(\log b)/2kA$.

3) Let $\sum_1^\infty (-1)^{n-1} a_n \equiv \sum_1^\infty (-1)^{n-1} \frac{\log D_n}{D_n}$. If $f(x) = \frac{x}{\log x}$, $g = \frac{1}{2k}$ and so

$$s' - s = \frac{1}{2k} \lim_{p,q\to\infty} \left[\frac{(\log x)^2}{2}\right]_{D_{2q}}^{D_{2p}}$$

if the limit exists.

Suppose that as $p > q \to \infty$,

$$\frac{D_{2p}}{D_{2q}} = 1 + \frac{a + o(1)}{\log D_{2q}}.$$

Then a simple calculation shows that the alteration in the sum of the series is $a/2k$.

3

These examples typify, in the order of their discussion, the three cases: $\lim_{x\to\infty} f(x)/x = \infty$, $0 < \lim f(x)/x < \infty$, $\lim f(x)/x = 0$. The function $f(x)$ which figures in them is such that $\int \frac{dx}{f(x)}$ can be readily evaluated. Under certain conditions, it is possible to find $s' - s$ without evaluating the integral. The conditions are set forth below in three theorems which reflect the differences between the three types of restriction on $f(x)$.

THEOREM 2. *If* $\lim_{n\to\infty} \frac{a_n \cdot D_n}{D_n - D_{n-2}} = g$, *then* $s' - s = g \lim_{p,q\to\infty} \log\left(\frac{D_{2p}}{D_{2q}}\right)$ *provided the limit exists.*

PROOF. Take $f(x) = x$ in (7)

COROLLARY. *Setting* $D_n = n, l_1 n, l_2 n \cdots$ *successively, we obtain Theorems* B *and* B'.

THEOREM 3. *If* $f(x)/x$ *tends steadily to* ∞ *with* x,

$$\lim_{n\to\infty} \frac{a_n \cdot f(D_n)}{D_n - D_{n-2}} = g, \qquad \frac{D_{2p} - D_{2q}}{f(D_{2q})} = O(1) \qquad (p > q \to \infty),$$

then s *is unaltered by the rearrangement.*[6]

[6] It is obvious that Theorems 3, 4 and all deductions therefrom can be suitably restated in the case $q > p$.

PROOF. Setting $f(x) = x\cdot\varphi(x)$, we have

$$\int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} = \frac{1}{\varphi\{D_{2q} + \theta(D_{2p} - D_{2q})\}} \int_{D_{2q}}^{D_{2p}} \frac{dx}{x} \qquad (0 < \theta < 1)$$

$$= \frac{\log\left\{1 + \dfrac{D_{2p} - D_{2q}}{f(D_{2q})}\varphi(D_{2q})\right\}}{\varphi\{D_{2q} + \theta(D_{2p} - D_{2q})\}} = o(1) \qquad (q \to \infty).$$

The desired result now follows from Theorem 1a.

COROLLARY. *The particular case $D_n = n$ contains Theorem* C.

More generally, we may take $D_n = l_1 n, l_2 n, \cdots$

THEOREM 4. *If, in Theorem* 1, $f(x)$ *satisfies the additional condition*

$$\lim_{x\to\infty} f'(x) = 0,$$

then $s' - s = 0$ *or* ag *or* ∞ *according as* $\lim\limits_{p>q\to\infty} \dfrac{D_{2p} - D_{2q}}{f(D_{2q})} = 0$ *or* a *or* ∞.

PROOF. Since

$$\int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} = \frac{D_{2p} - D_{2q}}{f(D_{2q}) + \theta(D_{2p} - D_{2q})f'\{D_{2q} + \eta(D_{2p} - D_{2q})\}} \qquad (0 < \eta < \theta < 1),$$

the result sought is an immediate consequence of (7).

REMARK. To bring out the relation between Theorems 3 and 4, we need only vary the form of the hypotheses in the latter to the extent of replacing the restriction on $f'(x)$ by the conditions: $f(x)$ tends steadily to ∞ with x and $f(x)/x$ tends steadily to zero. These conditions imply that $\lim f'(x) = 0$ and therefore, our previous proof of Theorem 4 is still valid.

COROLLARY 1. *When $D_n = n$, $a_n = 1/f(n)$ in the special form of Theorem* 4 *noticed above, we have Theorem* A.

COROLLARY 2. Let $D_n = l_x n$ $(x \geqq 1)$,

$$f(x) = (l_k x)^{m_k}(l_{k+1}x)^{m_{k+1}} \cdots (l_\lambda x)^{m_\lambda}\xi(x) \equiv \psi(x) \text{ (say)}$$

where $k \geqq 1$, the *ms* are positive or negative numbers, $\xi(x)$ has a continuous derivative and is such that $\xi(x) \to \infty$, $\xi'(x) = 0(1/x)$ $(x \to \infty)$. Theorem 4 then gives the following equivalent of (38a) and (38b) in Pringsheim's paper:

If $\lim_{n\to\infty} l_0 n\cdot l_1 n\cdot l_2 n \cdots l_{x-1}n\cdot\psi(l_x n)\cdot a_n = g$ $(l_0 n = n, x \geqq 1)$, *then* $s' - s = 0$ *or* $ag/2$ *or* ∞ *according as* $\lim\limits_{p>q\to\infty} \dfrac{l_x(2p) - l_x(2q)}{\psi(l_x 2q)} = 0$ *or* a *or* ∞.

COROLLARY 3. Let $D_n = l_x n$ $(x \geqq 1)$, $f(x) = x^{1-\rho}\cdot\psi(x)$ $(\rho > 0)$. We have then the result:

If $\lim_{n\to\infty} l_0 n\cdot l_1 n\cdot l_2 n \cdots l_{x-1}n\cdot(l_x n)^{1-\rho}\cdot\psi(l_x n)\cdot a_n = g$ $(l_0 n = n, x \geqq 1)$, *then* $s' - s = 0$ *or* $ag/2$ *or* ∞ *according as* $\lim\limits_{p>q\to\infty} \dfrac{l_x(2p) - l_x(2q)}{(l_x 2q)^{1-\rho}\cdot\psi(l_x 2q)} = 0$ *or* a *or* ∞.

This is equivalent to (44) in Pringsheim's paper.

NOTE. Theorem 4 thus includes two of the limiting cases discussed by Pringsheim: $\lim na_n = \infty$ as in Corollary 1, $\lim na_n = 0$ as in Corollaries 2 and 3.

3.1. I conclude with some applications of the theorems in §3. It is assumed in all these applications that $d_n \sim k$ as in §2.2.

1) If the series $\sum_1^\infty (-1)^{n-1}/D_n \cdot \log D_n$ is rearranged so that $D_{2p} = O(D_{2q} \cdot \log D_{2q})$ $(p > q \to \infty)$, then its sum is unaffected by the rearrangement (Theorem 3).

2) If the series $\sum_1^\infty (-1)^{n-1}/\log D_n$ is rearranged so that $D_{2p} - D_{2q} \sim a \cdot \log D_{2q}$ $(p > q \to \infty)$, then the alteration in its sum is $a/2k$ (Theorem 4).

3) Consider the series in (9), introducing the additional restriction $d_n \sim k$. (11) then gives

$$\lim \frac{a_n \cdot D_n^{\gamma+\delta-\alpha-\beta}}{D_n - D_{n-2}} = \frac{1}{2k} \frac{\Pi_{(d_n)}(\gamma) \cdot \Pi_{(d_n)}(\delta)}{\Pi_{(d_n)}(\alpha) \cdot \Pi_{(d_n)}(\beta)};$$

so that if $0 < \gamma + \delta - \alpha - \beta < 1$, a_n is subject to the hypothesis in Theorem 4, with $f(x) = x^{\gamma+\delta-\alpha-\beta}$; and if $\gamma + \delta - \alpha - \beta = 1$, a_n is as in Theorem 2. Hence we are led to formulate:

Suppose that for the series in (9) $d_n \sim k$.

(i) *If* $0 < \gamma + \delta - \alpha - \beta < 1$ *and if the series is rearranged so that* $D_{2p} - D_{2q} \sim aD_{2q}^{\gamma+\delta-\alpha-\beta}$ $(p > q \to \infty)$, *then the alteration in the sum of the series due to the rearrangement is*

$$\frac{a}{2k} \frac{\Pi_{(d_n)}(\gamma) \cdot \Pi_{(d_n)}(\delta)}{\Pi_{(d_n)}(\alpha) \cdot \Pi_{(d_n)}(\beta)}.$$

(ii) *If* $\gamma + \delta - \alpha - \beta = 1$ *and if in the rearrangement* $D_{2p} \sim bD_{2q}$ $(p > q \to \infty)$, *then the alteration is* $\dfrac{\log b}{2k} \cdot \dfrac{\Pi_{(d_n)}(\gamma) \cdot \Pi_{(d_n)}(\delta)}{\Pi_{(d_n)}(\alpha) \cdot \Pi_{(d_n)}(\beta)}$.

In the particular case $d_n = 1$, we have:

Suppose that the series

$$\sum_1^\infty (-1)^{n-1} \frac{(\alpha + 1)(\alpha + 2) \cdots (\alpha + n) \cdot (\beta + 1)(\beta + 2) \cdots (\beta + n)}{(\gamma + 1)(\gamma + 2) \cdots (\gamma + n) \cdot (\delta + 1)(\delta + 2) \cdots (\delta + n)}$$

where $\alpha, \beta, \gamma, \delta$ *are real and* $\gamma, \delta \neq -1, -2, \cdots$, *is rearranged* (i) *in the case* $0 < \gamma + \delta - \alpha - \beta < 1$, *so that* $p - q \sim aq^{\gamma+\delta-\alpha-\beta}$ $(p > q \to \infty)$, (ii) *in the case* $\gamma + \delta - \alpha - \beta = 1$, *so that* $p \sim bq$ $(p, q \to \infty)$. *Then the alterations in the sum of the series in* (i) *and* (ii) *are respectively* $\dfrac{a}{2^{\gamma+\delta-\alpha-\beta}} \dfrac{\Gamma(1 + \gamma) \cdot \Gamma(1 + \delta)}{\Gamma(1 + \alpha) \cdot \Gamma(1 + \beta)}$ *and* $\dfrac{\log b}{2} \dfrac{\Gamma(1 + \gamma) \cdot \Gamma(1 + \delta)}{\Gamma(1 + \alpha) \cdot \Gamma(1 + \beta)}$.

This result in a somewhat less general form has been proved from first principles by B. C. Das [1].

MADRAS CHRISTIAN COLLEGE.
TAMBARAM (MADRAS).
S. INDIA.

References

[1] Das, B. C., Effect of derangement of terms on some semi-convergent series, *Math. Student*, 5 (1937), 12–18.

[2] Littlewood, J. E., Note on the convergence of series of positive terms, *Messenger of Maths.*, 39 (1910), 191–192.

[3] Pringsheim, A., Allgemeine Theorie der Divergenz und Convergenz von Reihen mit Positiven Gliedern, *Math. Annalen*, 35 (1890), 297-394.

[4] Pringsheim, A., Zur Theorie der Dirichlet'schen Reihen, *Math. Annalen*, 37 (1890), 38-60.

[5] Pringsheim, A., Ueber die Werthveränderungen bedingt convergenter Reihen und Producte, *Math. Annalen*, 22 (1883), 455-503.

[6] Rajagopal, C. T., Convergence theorems for series of positive terms, *Journ. Indian Math. Soc.* (New Series), 3 (1938), 118–125.

[7] Rajagopal, C. T., On certain theorems of Pringsheim, *Tôhoku Math. Journ.*, 43 (1937), 122–126.

[8] Rajagopal, C. T., Some theorems connected with Maclaurin's integral test, *Math. Gazette*, 23 (1939), 456–461.

Annals of Mathematics
Vol. 42, No. 3, July, 1941

ON MEAN ONE-VALENT FUNCTIONS

By D. C. Spencer[1]

(Received October 20, 1939)

1. In several recent papers[2] I have shown that many of the properties of p-valent functions are possessed by the wider family of functions which are p-valent on the average. The present paper is a study of the finer structure of functions one-valent on the average in which it is shown that such functions possess some of the more striking properties of schlicht functions. The method of the paper also yields two results of a more general nature which are stated in §'s (7) and (8) below (Theorems 3 and 4). The main results have been stated without proof elsewhere.[3]

2. For completeness we begin by restating the definition of mean p-valency. We suppose that $f(z)$ is regular in $|z| < 1$, and that W is the Riemann domain which is the transform of $|z| < 1$ by $f(z)$. Let $W(R)$ be the area (multiply covered regions being counted multiply) of that portion of W which lies in the circle $|w| \leqq R$. Then if

$$W(R) \leqq \mathrm{p}\pi R^2 \tag{2.1}$$

for all $R > 0$, we say that $f(z)$ is mean p-valent.

It is convenient to express the inequality (2.1) analytically. Let $n(r, w)$ be the number of times (necessarily bounded by some number depending upon r) that $f(z)$ takes the value w in $|z| < r$. We define

$$p(r, R) = \frac{1}{2\pi}\int_{-\pi}^{\pi} n(r, Re^{i\psi})\, d\psi; \tag{2.2}$$

$$n(w) = \lim_{r\to 1} n(r, w); \qquad p(R) = p(1, R) = \lim_{r\to 1} p(r, R). \tag{2.3}$$

Then

$$\begin{aligned} W(R) &= \lim_{r\to 1}\int_0^R \int_{-\pi}^{\pi} n(r, Re^{i\psi}) R\, dR\, d\psi \\ &= \int_0^R \left(\lim_{r\to 1}\frac{1}{2\pi}\int_{-\pi}^{\pi} n(r, Re^{i\psi})\, d\psi\right) d(\pi R^2) = \int_0^R p(R)\, d(\pi R^2), \end{aligned} \tag{2.4}$$

---

[1] I am indebted to Mr. K. J. Arnold for making the drawing which is reproduced in Fig. 4 below, on the original of which 679 values of the function $K(z) = 4z/(1+z)^2$ were plotted.

[2] For example, references (2) and (3) at the end of the paper.

[3] Spencer (5).

and so (2.1) may be written in the form

$$\int_0^R p(R)\, d(\pi R^2) \leqq \mathrm{p}\pi R^2 \qquad (R > 0). \tag{2.5}$$

If we omit the averaging at (2.2), and suppose that

$$\int_0^R n(Re^{i\psi})\, d(\pi R^2) \leqq \mathrm{p}\pi R^2 \tag{2.6}$$

for all $R > 0$, uniformly in ψ, then we say that f is strongly mean p-valent.[4] Fig. 1 represents a domain W which is strongly mean 1-valent (*a fortiori* mean 1-valent), but which is not schlicht.

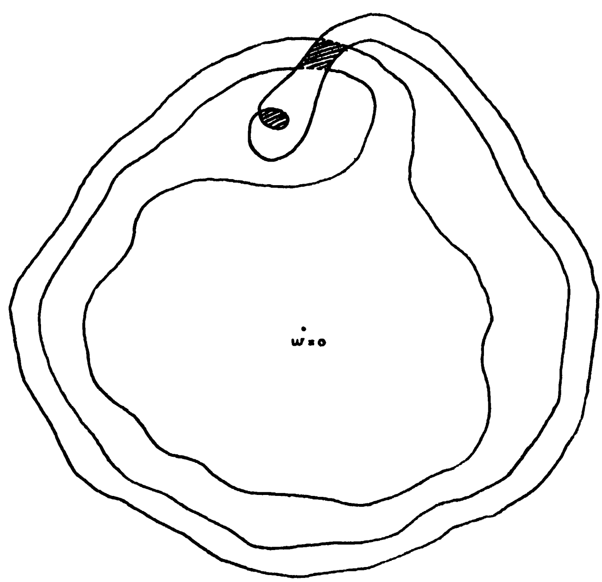


FIG. 1

3. It will be convenient to sum up here for future reference certain trivial deductions from mean 1-valency.

LEMMA 1. *Suppose that*

$$f(z) = a_1 z + a_2 z^2 + \cdots$$

is mean 1-*valent, and let*

$$d = \operatorname*{Inf}_{|z|<1} \left| \frac{f(z)}{z} \right|.$$

Then $d > 0$, *and* $n(w) = 1$ *if* $0 \leqq |w| < d$. *Furthermore, there exists a real number* β *such that* $n(de^{i\beta}) = 0$.

[4] Strong mean p-valency is therefore roughly midway in strength between p-valency and mean p-valency.

In fact, since $f(0) = 0$, there exists a circle of radius $\delta > 0$, and center $w = 0$, every point P of which is covered at least once by W. Since, by the definition of mean 1-valency, it is covered on the average at most once, it follows that the circle is covered exactly once by W. Hence d, the distance of the boundary of W from $w = 0$, satisfies $d \geqq \delta > 0$.

Suppose now that

$$n(de^{i\beta}) > 0 \tag{3.1}$$

for every β. To every β, $-\pi \leqq \beta < \pi$, there then corresponds a circle of radius $\rho = \rho(\beta)$ and center $de^{i\beta}$ every point of which is covered at least once, and it follows from the Heine-Borel theorem that there is a number $\epsilon > 0$ such that $n(w) \geqq 1$ for $|w| < d + \epsilon$. But there is then at least one point w_0, $|w_0| < d + \epsilon$, for which

$$n(w_0) > 1. \tag{3.2}$$

Otherwise the intersection of W with the circle $|w| < d + \epsilon$ would be schlicht, and its boundary points in $|w| < d + \epsilon$ would therefore be uncovered, contradicting (3.1). (3.2) implies that some neighborhood of w_0 is covered at least twice, and, since every point of the circle $|w| < d + \epsilon$ is covered at least once, we see that

$$W(d + \epsilon) = \int_0^{d+\epsilon} p(R)\, d(\pi R^2) > \pi(d + \epsilon)^2.$$

This is impossible by the hypothesis of mean 1-valency, and hence (3.1) is false.

4. We come now to the first result:

THEOREM 1. *Suppose that*

$$f(z) = a_1 z + a_2 z^2 + \cdots$$

is mean 1-*valent. Then*

$$|a_2| \leqq 2|a_1|, \tag{4.1}$$

and equality occurs if, and only if, $f(z) = a_1 z/(1 + \gamma z)^2$, $|\gamma| = 1$.

We write $z = re^{i\varphi}$, $w = f(z) = Re^{i\Phi}$. Then Theorem 1 follows quite easily from the following lemma, which is a generalization of the Bierberbach-Faber *Flächensatz*.

LEMMA 2. *If* $f(z)$ *is mean* 1-*valent and if* $\alpha > 0$, *then*

$$\int_{|z|=r} R^{-\alpha}\, d\Phi \geqq 0 \qquad (r < 1). \tag{4.2}$$

Lemma 2 when f is schlicht has been given by Prawitz,[5] who has used it to prove (4.1) for schlicht functions. The deduction of (4.1) from Lemma 2 is the same in our case, and I omit the calculations.

[5] See reference (2) at the end of the paper.

It remains, then, to prove Lemma 2. We suppose that $0 < r < 1$. Let $W(r)$ be the transform of $|z| < r$ by f, $B(r)$ its boundary (so that $B(r)$ is the transform of $|z| = r$); and let $d(r)$ (see Fig. 2)[6] be the distance of $B(r)$ from $w = 0$. Since the function $f(rz)$ (with picture $W(r)$) is *a fortiori* mean 1-valent

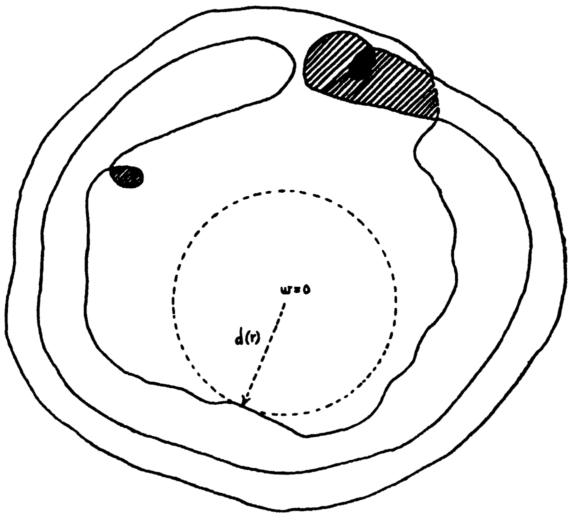


FIG. 2

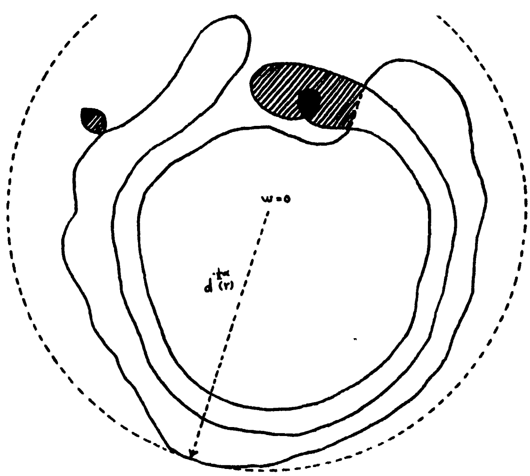


FIG. 3

and since $f(0) = 0$, $d(r) > 0$ by Lemma 1. Let $P = Re^{i\Phi}$ be a point of $W(r)$, and let $W'(r)$ (see Fig. 3) be the domain composed of the points $P' = R^{-\frac{1}{2}\alpha}e^{i\Phi}$.

[6] Incidentally, Fig. 2 represents a W which is mean 1-valent without being strongly mean 1-valent.

Then $W'(r)$ contains the point $w = \infty$, but is bounded ("internally") by a closed (but not necessarily simple) curve $B'(r)$ lying in the circle $|w| \leqq [d(r)]^{-\frac{1}{2}\alpha}$. Let $W'(r, R)$ be the area (multiply covered areas being counted multiply) of that portion of $W'(r)$ which lies in the circle $|w| < R$. Then

$$\frac{1}{2}\int_{|z|=r} R^{-\alpha}\, d\Phi = \pi d^{-\alpha}(r) - W'(r, d^{-\frac{1}{2}\alpha}(r)), \tag{4.3}$$

and the lemma asserts that this is non-negative.

We have, expressing (4.3) in terms of $p(R)$,

$$\frac{1}{2}\int_{|z|=r} R^{-\alpha}\, d\Phi = \int_0^{d^{-\frac{1}{2}\alpha}} (1 - p(r, R^{-2/\alpha}))\, d(\pi R^2)$$

$$= 2\pi \int_{d^{-\frac{1}{2}\alpha}(r)}^{\infty} (1 - p(r, R^{2/\alpha}))\, \frac{dR}{R^3},$$

and so, dropping the parameter r from our notation,

$$= \pi\alpha \int_d^{\infty} (1 - p(R))\, \frac{dR}{R^{1+\alpha}} = \pi\, d^{-\alpha} - \pi\alpha \int_d^{\infty} p(R)\, \frac{dR}{R^{1+\alpha}}. \tag{4.4}$$

Integrating by parts with respect to $p(R)$, and observing that the integrated part vanishes at the limits d and ∞, we have

$$\int_d^{\infty} p(R)\, \frac{dR}{R^{1+\alpha}} = (1 + \alpha) \int_d^{\infty} \left(\int_d^{R_1} p(R)\, dR\right) \frac{dR_1}{R_1^{2+\alpha}}.$$

Since, by Lemma 1, $p(R) = 1$ for $R < d$, we have

$$\int_d^{R_1} p(R)\, dR = \int_0^{R_1} p(R)\, dR - d.$$

But, integrating by parts,

$$\int_0^{R_1} p(R)\, dR = \int_0^{R_1} p(R) R \cdot \frac{dR}{R} = \frac{1}{R_1}\int_0^{R_1} p(R) R\, dR + \int_0^{R_1} \left(\int_0^{R} p(\underline{R})\underline{R}\, d\underline{R}\right)\frac{dR}{R^2}$$

$$\leqq \frac{R_1}{2} + \frac{1}{2}\int_0^{R_1} dR_1 = R_1$$

by the hypothesis of mean 1-valency. Hence

$$\int_d^{R_1} p(R)\, dR = \int_0^{R_1} p(R)\, dR - d \leqq R_1 - d,$$

and so

$$\int_d^{\infty} p(R)\, \frac{dR}{R^{1+\alpha}} \leqq (1 + \alpha) \int_d^{\infty} (R_1 - d)\, \frac{dR_1}{R_1^{2+\alpha}} = \frac{1 + \alpha}{\alpha} d^{-\alpha} - d^{-\alpha} = \frac{1}{\alpha} d^{-\alpha}. \tag{4.5}$$

Substituting from (4.5) into (4.4), we obtain

$$\frac{1}{2}\int_{|z|=r} R^{-\alpha}\, d\Phi \geqq \pi d^{-\alpha}(r) - \pi d^{-\alpha}(r) = 0,$$

and this proves Lemma 2.

By a suitable choice of α in Lemma 2 it can be shown (as Prawitz has pointed out in the case of schlicht functions) that

$$| a_3 | \leqq 3.05 | a_1 |.$$

I have not attempted to prove that $| a_3 | \leqq 3 | a_1 |$ (with equality only for the Koebe function), but it seems to me that this result must certainly be true (though possibly difficult to prove).

5. The two most important properties of 1-valency (more generally of p-valency) are: (i) an invariance to one-valent transformations; and (ii) the metrical property that large areas cannot accumulate in small neighborhoods.[7] Of these mean 1-valency (more generally mean p-valency) possesses only the second; the group property (i) is lacking. In fact, mean p-valency is not even an invariant to translations, the simplest transformations of all. When f is 1-valent (schlicht), it is possible by use of the property (i) to deduce immediately from Theorem 1 that

$$d \geqq \tfrac{1}{4} | a_1 |, \tag{5.1}$$

where d is the distance of the boundary of the map W from $w = 0$; but when f is only mean 1-valent the argument breaks down completely. The result is that I am able to prove (5.1) only for strongly mean 1-valent functions, and the precise inequality for mean 1-valent functions remains undetermined. *However, under mean 1-valency the weaker result is true that* $d \geqq | a_1 |/A$, *where* A *is an absolute constant less than* 7; and more detailed computations indicate (but in their present form do not prove) that the result is still true when A is only slightly greater than 4. A discussion of this problem is to be found at the end of the paper.

6. We now take up the result (5.1) for strongly mean 1-valent functions. Afterwards we state and prove the two general theorems mentioned in §1 (Theorems 3 and 4 below), the hypotheses of which are essentially the conclusion of Lemma 1 (here only a vestige of mean 1-valency appears).

THEOREM 2. *If $f(z)$ is strongly mean 1-valent, then*

$$d \geqq \tfrac{1}{4} | a_1 |$$

with equality if, and only if, $f(z) = a_1 z/(1 + \gamma z)^2$, $| \gamma | = 1$.

The following result is almost an immediate consequence of Theorem 2:

COROLLARY. *If $f(z)$ is strongly mean 1-valent, then*

$$| f(z) | \geqq \frac{| a_1 | \cdot | z |}{(1 + | z |)^2},$$

and there is equality if, and only if,

$$f(z) = a_1 z/(1 + \gamma z)^2, \qquad | \gamma | = 1.$$

[7] It is the distribution of the area of the map with respect to $w = 0$ which determines the magnitude of the mean values of $| f |$ and $| f' |$ (and so of the upper bounds of the coefficients).

In fact, more generally, the inequality

$$|f(z)| \geqq \frac{4}{A} \cdot \frac{|a_1| \cdot |z|}{(1 + |z|)^2}$$

is a consequence of $d \geqq |a_1|/A$ for *any* f. I omit a proof of this statement since the proof is substantially the same as in the well-known case when f is schlicht (see R. Nevanlinna, *Eindentige analytische Funktionen*, Berlin, 1936).

In the sequel we shall make repeated use of the following lemma. Suppose that

$$g(z) = b_1 z + b_2 z^2 + \cdots$$

is regular in $|z| < 1$. Let G be the map of the unit circle by g, and let G^* be the "star" of G with respect to the point $g(0)$.[8] We write $Z^* = g^{-1}(G^*)$, and denote by $p_1^*(R)$, $p_2^*(R)$ the valency functions of G^* and Z^* respectively.[9]

LEMMA 3.

$$\lg |b_1| < \int_0^\infty \lg R\, d[-p_1^*(R)] + \int_0^\infty \lg R\, d[-p_2^*(R)] \tag{6.1}$$

unless $g = b_1 z$.

(6.1), in different notation, is an inequality of Bermant.[10]

The lemma remains true if g is regular only in Z^*, *but is false in general if this condition is not satisfied.*

To prove Theorem 2 we suppose that $d = 1$, and prove that then $|a_1| \leqq 4$, a statement which is plainly equivalent to the theorem. By lemma 1 there exists a β such that $f \neq e^{i\beta}$ in $|z| < 1$. We write

$$K_\beta(z) = 4e^{i\beta} z/(1 + z)^2.$$

This function (Koebe's function) maps the circle $|z| < 1$ on the plane slit along a straight line from $e^{i\beta}$ to ∞, and the inverse function $K_\beta^{-1}(w)$ has branch points of order 2 at $w = e^{i\beta}$ and at ∞. Since $f \neq e^{i\beta}$, ∞, however, either branch of K_β^{-1} is regular and one-valued on W, the map of $|z| < 1$ by f. To fix ideas we suppose that $K_\beta^{-1}(w)$ is the continuation over W of the branch which vanishes at $w = 0$,[11] and write

$$g(z) = K_\beta^{-1}\{f(z)\} = b_1 z + \cdots,$$

where $|b_1| = |a_1|/4$. By Lemma 3 we have

$$\lg \frac{|a_1|}{4} \leqq \int_0^\infty \lg R\, d[-p_1^*(R)] + \int_0^\infty \lg R\, d[-p_2^*(R)], \tag{6.2}$$

[8] More precisely, G^* is the subdomain of G any point of which can be reached by starting at $g(0)$ and travelling outward along some radius vector. G^* does *not* depend on the particular way in which the sheets of G are supposed stuck together.

[9] $p^*(R)$, for example, is therefore the average valency of G^* on the circumference $|g| = R$. Both G^* and Z^* are schlicht, and so p^* and p_2^* are bounded above by unity.

[10] For a proof of (6.1) in the above form see Spencer (5).

[11] We might equally well take the other branch (which is infinite at the origin).

and it is sufficient to show that

$$\int_0^\infty \lg R\, d[-p_1^*(R)] \leqq 0 \tag{6.3}$$

$$\int_0^\infty \lg R\, d[-p_2^*(R)] \leqq 0. \tag{6.4}$$

(6.4) is immediate. For Z^* is a schlicht domain which contains the point $z = 0$, and lies inside $|z| < 1$. Therefore by a known inequality[12] we have

$$\exp\left\{\int_0^\infty \lg R^2\, d[-p_2^*(R)]\right\} \leqq \int_0^\infty R^2\, d[-p_2^*(R)] = \frac{1}{\pi}\int_0^\infty p_2^*(R)\, d(\pi R^2) \leqq 1,$$

and this is equivalent to (6.4).

As for (6.3), we prove the following stronger result that

$$\int_0^\infty \lg R\, d[-p_1(R)] \leqq 0, \tag{6.5}$$

where $p_1(R)$ is the valency function of G.

Let $\mathfrak{E}(R_0)$ be the set of points in $|z| < 1$ at which $|g| = |K_\beta^{-1}\{f\}| > R_0$, and write

$$A(\beta, R_0) = \iint\limits_{\mathfrak{E}(R_0)} \left|\frac{d}{dz}\lg g\right|^2 r\, dr\, d\varphi = \iint\limits_{\mathfrak{E}(R_0)} \frac{|f'(re^{i\varphi})|^2}{|f(re^{i\varphi}) - e^{i\beta}|\cdot|f(re^{i\varphi})|^2}\, r\, dr\, d\varphi.$$

We suppose that R_0 is small enough to satisfy any condition imposed by the argument. Then $p_1(R) = 1$ for $R \leqq R_0$ by Lemma 1, and since $p_1(\infty) = 0$ we have, integrating by parts,

$$\begin{aligned} 2\pi\int_0^\infty \lg R\, d[-p_1(R)] &= 2\pi\int_{R_0}^\infty \lg R\, d[-p_1(R)] \\ &= \int_{R_0}^\infty p_1(R)\, d(2\pi \lg R) - 2\pi \lg\frac{1}{R_0} = A(\beta, R_0) - 2\pi \lg\frac{1}{R_0}. \end{aligned} \tag{6.6}$$

Next,

$$\begin{aligned} &\left|\iint\limits_{\mathfrak{E}(R_0)} \frac{|f'|^2}{|f - e^{i\beta}|\cdot|f|^2}\, r\, dr\, d\varphi - \underset{|f|>4R_0}{\int_0^1\int_{-\pi}^{\pi}} \frac{|f'|^2}{|f - e^{i\beta}|\cdot|f|^2}\, r\, dr\, d\varphi\right| \\ &\qquad \leqq \frac{C}{R_0^2} \underset{4R_0 - CR_0^2 < |f| < 4R_0 + CR_0^2}{\int_0^1\int_{-\pi}^{\pi}} |f'|^2 r\, dr\, d\varphi = \frac{C}{R_0^2}\int_{4R_0 - CR_0^2}^{4R_0 + CR_0^2} p(R)\, d(\pi R^2) \\ &\qquad = \frac{C}{R_0^2}\int_{4R_0 - CR_0^2}^{4R_0 + CR_0^2} d(\pi R^2) = O(R_0), \end{aligned}$$

by Lemma 1 and the hypothesis that $d = 1$; and hence

$$A(\beta, R_0) = \underset{|f|>4R_0}{\int_0^1\int_{-\pi}^{\pi}} \frac{|f'|^2}{|f - e^{i\beta}|\cdot|f|^2}\, r\, dr\, d\varphi + O(R_0). \tag{6.7}$$

---

[12] See Spencer (5).

But

$$\int_0^1 \int_{-\pi}^{\pi} \frac{|f'|^2}{|f - e^{i\beta}|\cdot|f|^2}\, r\, dr\, d\varphi = \int_{4R_0}^{\infty} \int_{-\pi}^{\pi} \frac{n(Re^{i\Phi})}{|e^{i\beta} - Re^{i\Phi}|\cdot R^2}\, R\, dR\, d\Phi$$

$$(6.8) \qquad = \int_{\substack{4R_0\\ |f|>4R_0}}^{\infty} \int_{-\pi}^{\pi} \frac{n(Re^{i\Phi})\, dR\, d\Phi}{R\sqrt{1 - 2R\cos(\Phi - \beta) + R^2}}$$

$$= \int_{4R_0}^{\infty} \int_{-\pi}^{\pi} \frac{n(Re^{i(\Phi+\beta)})}{R\sqrt{1 - 2R\cos\Phi + R^2}}\, dR\, d\Phi.$$

Substituting from (6.8) into (6.7) we obtain the formula (valid for all sufficiently small R_0):

$$(6.9) \qquad A(\beta, R_0) = \int_{4R_0}^{\infty} \int_{-\pi}^{\pi} \frac{n(Re^{i(\Phi+\beta)})}{R\cdot\sqrt{1 - 2R\cos\Phi + R^2}}\, dR\, d\Phi + O(R_0).$$

Integrating by parts we have

$$\int_{4R_0}^{\infty} \frac{n(Re^{i(\Phi+\beta)})}{R\cdot\sqrt{1 - 2R\cos\Phi + R^2}}\, dR = \left[\frac{1}{\sqrt{1 - 2R\cos\Phi + R^2}} \int_{4R_0}^{R} \frac{n}{\underline{R}}\, d\underline{R}\right]_{4R_0}^{\infty}$$

$$- \int_{4R_0}^{\infty} \left(\int_{4R_0}^{R} \frac{n}{\underline{R}}\, d\underline{R}\right) \frac{d}{dR}\left\{\frac{1}{\sqrt{1 - 2R\cos\Phi + R^2}}\right\} dR$$

$$(6.10) \qquad = -\int_{4R_0}^{\infty} \left(\int_{4R_0}^{R} \frac{n}{\underline{R}}\, d\underline{R}\right) \frac{d}{dR}\left\{\frac{1}{\sqrt{1 - 2R\cos\Phi + R^2}}\right\} dR$$

$$= \int_{4R_0}^{1} \left(\int_{4R_0}^{R} \frac{1}{\underline{R}}\, d\underline{R}\right) \frac{R - \cos\Phi}{(1 - 2R\cos\Phi + R^2)^{3/2}}\, dR$$

$$+ \int_{1}^{\infty} \left(\int_{4R_0}^{R} \frac{n}{\underline{R}}\, d\underline{R}\right) \frac{R - \cos\Phi}{(1 - 2R\cos\Phi + R^2)^{3/2}}\, dR,$$

since (by hypothesis) $d = 1$ and so, by Lemma 1, $n = 1$ for $0 \leqq R < 1$. We write

$$(6.11) \qquad I(R_0, \Phi) = \int_{4R_0}^{1} \left(\int_{4R_0}^{R} \frac{1}{\underline{R}}\, d\underline{R}\right) \frac{R - \cos\Phi}{(1 - 2R\cos\Phi + R^2)^{3/2}}\, dR.$$

Then $I(R_0, \Phi)$ depends only on R_0 and Φ, but *not* on f. Next, writing

$$(6.12) \qquad N(R, \Phi + \beta) = \int_{4R_0}^{R} \frac{n(\underline{R}e^{i(\Phi+\beta)})}{\underline{R}}\, d\underline{R},$$

we have

$$N(R, \Phi + \beta) = \left[\frac{1}{R}\int_{4R_0}^{R} n\, d\underline{R}\right]_{4R_0}^{R} + \int_{4R_0}^{R} \left(\int_{4R_0}^{R} n\, d\underline{R}\right) \frac{dR_1}{R_1^2}$$

$$\leqq \frac{1}{R}(R - 4R_0) + \int_{4R_0}^{R} (R_1 - 4R_0)\frac{dR_1}{R_1^2} = \lg \frac{R}{4R_0}$$

by the hypothesis of strong mean 1-valency and the fact that $n = 1$ if $R < 1$ (since $d = 1$). Hence

$$N(R, \Phi + \beta) \leqq \int_{4R_0}^{R} \frac{dR}{R} = \lg \frac{R}{4R_0}, \tag{6.13}$$

uniformly in Φ. Since

$$\frac{R - \cos \Phi}{(R^2 - 2R \cos \Phi + 1)^{3/2}} > 0$$

if $R > 1$, we have, by (6.10), (6.11), (6.12), and (6.13), that

$$\int_{4R_0}^{\infty} \frac{n(Re^{i(\Phi+\beta)})}{R \cdot \sqrt{1 - 2R \cos \Phi + R^2}} \, dR \leqq \int_{4R_0}^{\infty} \frac{1}{R \cdot \sqrt{1 - 2R \cos \Phi + R^2}} \, dR. \tag{6.14}$$

Integrating both sides of (6.14) with respect to Φ from $-\pi$ to π, we obtain the inequality

$$\begin{aligned} A(\beta, R_0) &= \int_{-\pi}^{\pi} \int_{4R_0}^{\infty} \frac{n(Re^{i(\Phi+\beta)})}{R \cdot \sqrt{1 - 2R \cos \Phi + R^2}} \, dR \, d\Phi + O(R_0) \\ &\leqq \int_{-\pi}^{\pi} \int_{4R_0}^{\infty} \frac{1}{R \cdot \sqrt{1 - 2R \cos \Phi + R^2}} \, dR \, d\Phi + O(R_0) \\ &= \int_{-\pi}^{\pi} \int_{R_0}^{1} \left| \frac{1}{z} \right|^2 r \, dr \, d\varphi + 0(R_0) = 2\pi \lg \frac{1}{R_0} + O(R_0), \end{aligned} \tag{6.15}$$

since when n is replaced by unity, we may suppose that $f = K_\beta$, and so $g = z$.

Finally, substituting from (6.15) into (6.6), we have that

$$\int_0^{\infty} \lg R \, d[-p_1(R)] \leqq O(R_0).$$

Letting $R_0 \to 0$, we obtain (6.5), and this completes the proof of Theorem 2 apart from the statement concerning equality. For this we note that equality implies by Lemma 3 that $g = z$, therefore that $f = K_\beta$.

7. Suppose that $f(0) = 0$. By Lemma 1 mean 1-valency then implies: (i) that $n(0) = 1$; and (ii) that $n(de^{i\beta}) = 0$ for at least one β. We now take for hypotheses only (i) and (ii), and obtain the following theorem:

THEOREM 3. *Suppose that*

$$f(z) = a_1 z + a_2 z^2 + \cdots$$

is regular for $|z| < 1$, *that* $n(0) = 1$, *and that*

$$d = \operatorname*{Inf}_{|z|<1} \left| \frac{f(z)}{z} \right|.$$

Then if there is a β *for which* $n(de^{i\beta}) = 0$, *we have*

$$d > \frac{1}{13} |a_1|. \tag{7.1}$$

The constant 13 is not the best possible. As a matter of fact, a better (and in a sense best possible) constant is obtainable by a "subordination" argument (Lindelöf's principle). But the theorem is only mildly interesting on its own account, and is included here mainly in order to deduce Theorem 4 below. The hypothesis that $n(de^{i\beta}) = 0$ for some β is necessary. This is seen by considering the function

$$f(z) = z - z^2,$$

for which $n(de^{i\beta}) = n(0) = 1$.

Since by hypothesis $n(0) = 1$, $n(de^{i\beta}) = 0$, we have $d > 0$. We suppose that $d = 1$, and prove that $|a_1| < 13$. Then if β is a number for which $n(e^{i\beta}) = 0$, the function[13] $g = K_\beta^{-1}\{f\}$ is regular for $|z| < 1$. By Lemma 3 we have, therefore,

$$\text{(7.2)} \qquad \lg \frac{|a_1|}{4} \leqq \int_0^\infty \lg R \, d[-p_1^*(R)] + \int_0^\infty \lg R \, d[-p_2^*(R)] \leqq \int_0^\infty \lg R \, d[-p_1^*(R)] = J^*, \text{ say,}$$

where $p_1^*(R)$ is the valency function of G^*. We find an upper bound for J^*.

Now every point of the circle $|w| < 1$ is covered exactly once. For $n(0) = 1$, and if some point w_0, $|w_0| < 1$, were covered n-times, where $n \neq 1$, then the projection of a boundary point would fall on the straight line connecting 0 with w_0; and this is impossible since $d = 1$.

We define

$$\text{(7.3)} \qquad t_\beta(\varphi) = \begin{cases} K_\beta\{\mathfrak{E}[\arg z = \varphi, 0 \leqq |z| < \infty]\}, & \varphi \neq 0, \pi \\ K_\beta\{\mathfrak{E}[\arg z = \varphi, 0 \leqq |z| < 1]\}, & \varphi = 0, \pi \end{cases}$$

where $\mathfrak{E}[\cdots]$ is the set of points defined by the conditions stated in the bracket. Positions lying outside $|w| < 1$ of the lines t_0 which cut $|w| = 1$ at intervals of 5° from 0° to 180° are shown in Fig. 4. We denote by W_β^* the subdomain of W any point P of which can be reached by starting at $w = 0$ and travelling outward along some line $t_\beta(\varphi)$, $\varphi = \varphi(P)$, always in the sense of increasing $|z|$. Then $G^* = K_\beta^{-1}\{W_\beta^*\}$. Also if we let $n_\beta^*(w)$ be the number of times W_β^* covers the point w, then

$$\text{(7.4)} \qquad n_\beta^*(w) \leqq 2$$

for all w.

Next, if R_0 is sufficiently small, we have[14]

$$\text{(7.5)} \quad 2\pi J^* = \int_{4R_0}^\infty \int_{-\pi}^{\pi} \frac{n_\beta^*(Re^{i(\Phi+\beta)})}{R \cdot \sqrt{1 - 2R\cos\Phi + R^2}} \, dR \, d\Phi - 2\pi \lg \frac{1}{R_0} + O(R_0).$$

That is to say, $2\pi J^*$ is the difference between the logarithmic area of G^* and

[13] The branch of K_β^{-1} which vanishes at $w = 0$ is again chosen.

[14] By an argument analogous to the one given in §6.

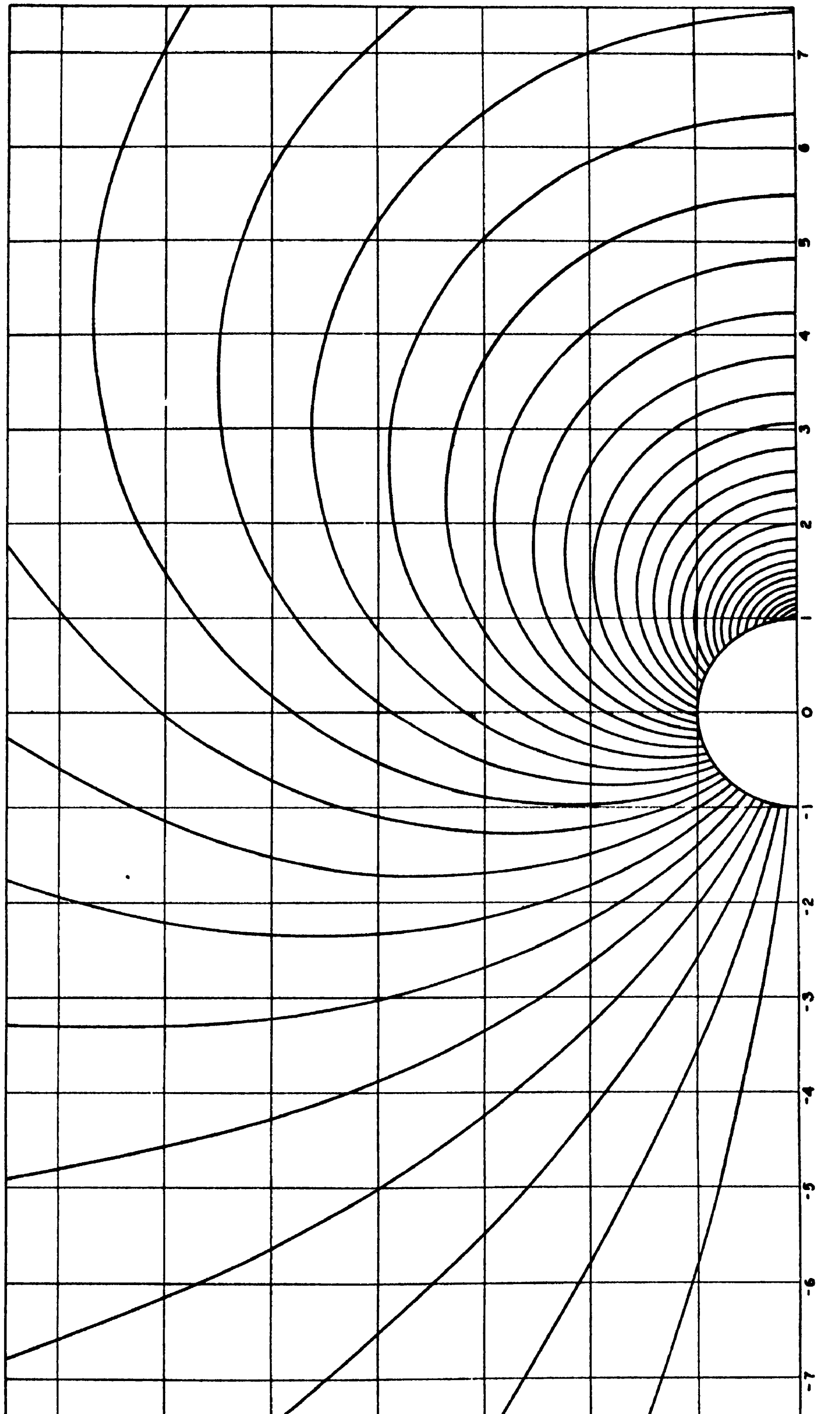


FIG. 4

that of the unit circle $|g| < 1$. If we combine (7.4) with the fact that $n_\beta^*(w) = n(w) = 1$ for $|w| < 1$, we see that

$$J^* \leqq J_0^*, \tag{7.6}$$

where J_0^* is obtained by taking

$$n_\beta^*(w) = \begin{cases} 1, \text{ if } |w| < 1 \\ 2, \text{ if } |w| > 1 \end{cases}$$

in (7.5)

Finally, we may express J_0^* in the form

$$J_0^* = \frac{1}{2\pi}\int_0^{2\pi} u\, dv(t),$$

where

$$u = \lg \{1 - 2\sqrt{\sin \tfrac{1}{2}t}\,(\cos \tfrac{1}{4}t + \sin \tfrac{1}{4}t) + 2 \sin \tfrac{1}{2}t\}.$$

$$v = t - 2\tan^{-1}\left\{\frac{\sqrt{\sin \tfrac{1}{2}t}\,(\cos \tfrac{1}{4}t - \sin \tfrac{1}{4}t}{1 - \sqrt{\sin \tfrac{1}{2}t}\,(\cos \tfrac{1}{4}t + \sin \tfrac{1}{4}t)}\right\}.$$

A computation gives

$$1.1541 < J_0^* < 1.1784, \tag{7.7}$$

and substituting the upper bound from (7.7) into (7.2), we get (7.1).

8. By combining Theorem 3 above with Theorem 3 of my paper (4)[15], we obtain the following theorem:

THEOREM 4. *Suppose that $f(z)$ satisfies the hypotheses of Theorem 3, and that, in addition, $|f(z)| < M$ in $|z| < 1$. Then, if $0 < \rho < 1$,*

$$|\arg f| \leqq B \lg \{1/(1 - \rho)\} + O(1), \tag{8.1}$$

where $B = \dfrac{1}{\pi} \lg (13M/|a_1|)$.

The constant in the $O(1)$ depends on M, $|a_1|$, and the particular determination of $\arg f$ chosen.

We write

$$f(z) = e^{u+iv}.$$

By Theorem 3 and the hypothesis that $|f| < M$, we have

$$\lg (|a_1|/13) \leqq u \leqq \lg M, \tag{8.2}$$

if $|v| > v_0$. We cut W (the map of $|z| < 1$ by f) from $w = 0$ along a radius to a boundary point $de^{i\beta}$, $n(de^{i\beta}) = 0$. We denote the resulting domain by W', and write $Z' = f^{-1}\{W'\}$. The function

$$h = \{f\}^{-i} = e^{v-iu}$$

[15] We shall refer to this paper as V_2.

is regular in Z', and we apply the method of V_2 to the function h. By (8.2) we have that

$$\lg \frac{|a_1|}{13} \leqq |\arg h| \leqq \lg M.$$

Hence, in the notation of V_2,

$$\Theta_\nu(R) \leqq R \lg \frac{13\, M}{|a_1|} \tag{8.3}$$

uniformly in ν, if $R > R_0 = e^{\nu_0}$. Substituting from (8.3) into Theorem 3 of V_2, we obtain (8.1).

Generalizations of Theorem 4 are possible. Condition (8.2) is much too strong; all that we require is that it is satisfied on the average. For example, in Theorem 4 it would be sufficient to suppose that $|f|$, *qua* function of arg f, is bounded when averaged with respect to arg f (as opposed to ordinary mean valency, which implies that $|\arg f|$ is bounded when averaged with respect to $|f|$).

9. Suppose that the projection of a Riemann domain W lies inside a schlicht domain D in the space of the variable w, and that T is regular and schlicht in D. We write

$$w' = R'e^{i\Phi'} = T(w), \qquad n'(w') = n(w), \qquad p'(R') = \frac{1}{2\pi}\int_{-\pi}^{\pi} n'(R'e^{i\Phi'})\, d\Phi'.$$

The lines R' = constant, Φ' = constant, define in D a new coördinate system. We shall say that W is mean **p**-valent (T) if for all $R' > 0$,

$$\int_0^{R'} p'(R')\, d(\pi R'^2) \leqq p\pi R'^2.$$

Suppose now that

$$f(z) = a_1 z + a_2 z^2 + \cdots$$

is regular in $|z| < 1$, that $f \neq e^{i\beta}$ in $|z| < 1$ (i.e., $n(e^{i\beta}) = 0$), and define W_β^* as in §7. We cut through all sheets of W_β^* along a radial line from $e^{i\beta}$ to ∞, and denote the resulting set of points by $\mathfrak{W}_\beta^*$. If there exists a β, $f \neq e^{i\beta}$ in $|z| < 1$, for which $\mathfrak{W}_\beta^*$ is mean 1-valent (K_β^{-1}), we say briefly that W is metrically subordinate to K. The following result is true:

If f is metrically subordinate to K, then

$$|a_1| \leqq 4, \tag{9.1}$$

with equality only if $f = K_\beta$.

In fact, let

$$G^* = K_\beta^{-1}\{W_\beta^*\}; \qquad \mathfrak{G}^* = K_\beta^{-1}\{\mathfrak{W}_\beta^*\}.$$

Then if $p_1^*(R)$, $\mathfrak{p}_1^*(R)$ are the valency functions of G^*, $\mathfrak{G}^*$ respectively, we have

$$(9.2)\qquad \mathfrak{p}_1^*(R) = \begin{cases} p_1^*(R) + p_1^*\left(\dfrac{1}{R}\right), & \text{if } 0 \leqq R < 1, \\ 0, \quad \text{if } R > 1. \end{cases}$$

Furthermore, by hypothesis,

$$(9.3)\qquad \int_0^R \mathfrak{p}_1^*(R)d(R^2) \leqq \begin{cases} R^2, & \text{if } R < 1 \\ 1, & \text{if } R > 1. \end{cases}$$

In particular, therefore, by Lemma 1, $\mathfrak{p}_1^*(R) = 1$ if $0 \leqq R < R_0$, and so (integrating by parts)

$$\int_0^\infty \lg R^2\, d[-\mathfrak{p}_1^*(R)] = \lg R_0^2 + \int_{R_0}^\infty \mathfrak{p}_1^*(R)\, \frac{d(R^2)}{R^2}$$

$$(9.4)\qquad = \lg R_0^2 + 2\int_{R_0}^\infty \left\{\int_{R_0}^R \mathfrak{p}_1^*(R)\, d(R^2)\right\}\frac{dR}{R^3}$$

$$= \lg R_0^2 + 2\int_{R_0}^1 \left\{\int_{R_0}^R \mathfrak{p}_1^*(R)d(R^2)\right\}\frac{dR}{R^3} + 2\int_1^\infty \left\{\int_{R_0}^R \mathfrak{p}_1^*(R)\, d(R^2)\right\}\frac{dR}{R^3}$$

$$= \lg R_0^2 + I_1 + I_2, \qquad \text{say,}$$

where by (9.3)

$$(9.5)\qquad I_1 = 2\int_{R_0}^1 \left\{\int_0^R \mathfrak{p}_1^*(R)\, d(R^2) - R_0^2\right\}\frac{dR}{R^3} \leqq -\lg R_0^2 + R_0^2 - 1$$

$$(9.6)\qquad I_2 \leqq 2\int_1^\infty \{1 - R_0^2\}\,\frac{dR}{R^3} = 1 - R_0^2.$$

Using (9.2) and substituting from (9.5) and (9.6) into (9.4), we obtain that

$$J^* = \int_0^\infty \lg R\, d[-p_1^*(R)] = \int_0^1 \lg R\, d[-p_1^*(R)] + \int_1^\infty \lg R\, d[-p_1^*(R)]$$

$$= \int_0^1 \lg R\, d[-p_1^*(R)] + \int_0^1 \lg R\, d\left[-p_1^*\left(\frac{1}{R}\right)\right] = \tfrac{1}{2}\int_0^\infty \lg R^2\, d[-\mathfrak{p}_1^*(R)] \leqq 0,$$

and this is equivalent, by Lemma 3, to (9.1).

10. The question now arises as to whether the result of Theorem 2 is true for mean 1-valent functions. I am not able to answer this question, the difficulty being as follows.

Let F_1 be the family of mean 1-valent functions, and let

$$J_1 = \operatorname*{Max}_{f \epsilon F_1} \int_0^\infty \lg R \, d[-p_1(R)], \qquad J_1^* = \operatorname*{Max}_{f \epsilon F_1} \int_0^\infty \lg R \, d[-p_1^*(R)],$$

where $p_1(R)$, $p_1^*(R)$ are the valency functions of G and G^* respectively (so that $J_1^* = \text{Max } J^*$). Then

$$(10.1) \qquad 0 < J_1^* < J_1 .$$

In fact, let W be the domain represented in Fig. 5, when $n(w) = 1$ for almost all w exterior to D_1 and D_2, and $n(w) = 2$ for $w \epsilon D_1$, $n(w) = 0$ for $w \epsilon D_2$. The solid curved lines are (we suppose) lines $t_0(\varphi)$. The boundary of W consists of the boundaries of D_1 and D_2, and slits. Let $D_1(R)$, $D_2(R)$ be the widths of D_1 and D_2 on $| w | = R$. We suppose: (i) that $D_1(R) = D_2(R)$ for all R; (ii) that

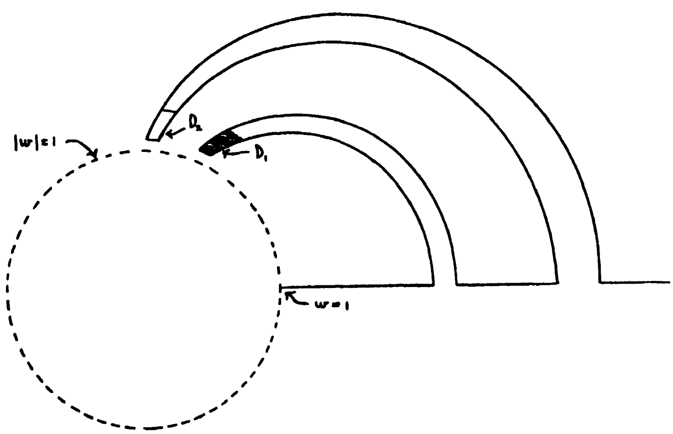


FIG. 5

all of W is swept out by lines $t_0(\varphi)$ as φ varies from 0 to 2π—that is to say, $W = W_0^*$. Both assumptions may plainly be fulfilled if D_1, D_2 have suitable shapes, and (i) implies that $p(R) = 1$ for all $R > 0$, so that W is mean 1-valent. Now write $w = Re^{i\Phi}$, $dW_w = R \, dR \, d\Phi$, and let dW_g be the transform of dW_w by $K_\beta^{-1}(w)$. Then if $\mathfrak{R}$ is the distance from the element dW_w to the point $e^{i\beta}$, we have by (6.8) with $\beta = 0$ that

$$dW_g = dW_w/\Lambda, \qquad \text{say,}$$

where $\Lambda = R^2\mathfrak{R}$. By sliding points $w = Re^{i\Phi}$ from one of the two sheets covering D_1 along the circumference $| w | = R$ into D_2, we can just fill the latter domain, and still have D_1 covered once. For the resulting picture $J^* = 0$. Since $\Lambda(Re^{i\Phi})$ is a *decreasing* function of Φ as we slide w along the circumference $| w | = R$ from D_1 into D_2, we see that the original J^* corresponding to Fig. 5 must be positive; and (since $J^* \leqq J_1^*$) this proves the first half of (10.1).

As for the second half, we may plainly suppose (by a trivial rotation of each function f) that $\beta = 0$. Then it is only necessary to point out that an argument similar to the above shows that J_1 is the area which results from the choice

$$n_1(Re^{i\Phi}) = \begin{cases} 1, & 0 \leqq R < 1 \\ 2, & R > 1, \quad -\frac{\pi}{2} < \Phi < \frac{\pi}{2}\,; \\ 0, & R > 1, \quad \frac{\pi}{2} \leqq \Phi \leqq \frac{3\pi}{2} \end{cases} \tag{10.2}$$

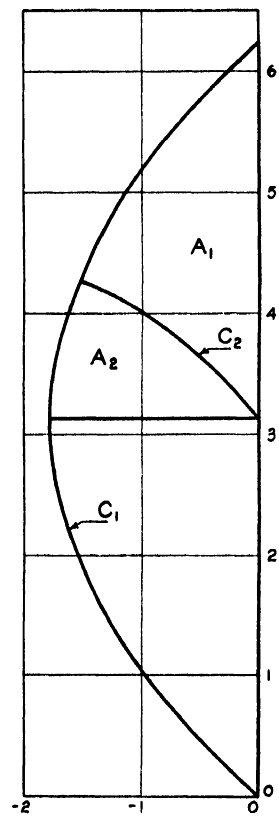


FIG. 6

In Fig. 6, C_1 is the transform of $|w| = 1$, C_2 the transform of any $w = 3\pi/2$ by $\lg K_0^{-1}(w)$, and $J_1 = \frac{1}{\pi}(A_1 - A_2)$. For any function n which satisfies mean 1-valency and differs from n_1 in a set of positive plane measure, the corresponding J is *less* than J_1. In particular, the function n^* corresponding to any W_0^*

plainly differs from n_1 in a set of positive plane measure;[16] and this proves the second half of (10.1).

Thus the method of this paper fails to yield the result

$$d \geqq \tfrac{1}{4}|a_1|. \tag{10.3}$$

for the family F_1 of mean 1-valent functions.

It is perhaps worthwhile to point out that the area integral J (also J^*) is a function of the parameter β, and that $J(\beta)$ exist for all β whether $n(e^{i\beta}) = 0$ or not. This can be seen from the formula (6.9). Furthermore,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} J(\beta)\,d\beta \leqq 0 \tag{10.4}$$

if $f \in F_1$. In fact, the integration has the effect of replacing n by p in the formulas of §6, and the argument given there (with strong mean 1-valency replaced by mean 1-valency) yields (10.4). From (10.4) we see that we can choose $a\beta$, β_0 say, for which

$$J(\beta_0) \leqq 0. \tag{10.5}$$

But we cannot deduce (10.3) from (10.5). The difficulty is that for this choice of β the function $g = K_\beta^{-1}\{f\}$ is not necessarily regular in Z^*, and we cannot, therefore, apply Lemma 3.

On the positive side it is true, of course, that

$$d \geqq |a_1|/A, \tag{10.6}$$

where A is an absolute constant. By Lemma 1 we see that a mean 1-valent f satisfies the hypotheses of Theorem 3, and therefore that $A < 13$. But a rather crude computation of J_1 shows that (10.6) is true with $A < 7$. Finally the inequality (10.1) indicates that the determination of J_1^* would still further reduce A. Although I have not attempted to determine J_1^*, I should like to conclude the paper by pointing out certain qualitative considerations which indicate (though of course they do not prove) that J_1^* (though positive) is small, and that (10.6) is therefore probably true with $A = 4$.

We suppose again (without loss of generality) that $\beta = 0$. The two lines $t_0(\varphi)$, $t_0(\varphi')$ which cut the circumference $|w| = 1$ at $\Phi - 1^0$ and at Φ respectively, $\Phi \geqq 1^0$, together with a portion of $|w| = 1$ and the real axis, bound a certain region of the w-plane which we shall denote by $\Delta(\Phi)$. We denote by $\alpha(\Phi)$ the area of the transform of $\Delta(\Phi)$ by $\lg K_0^{-1}(w)$. The function $\alpha(\Phi)$, labelled "area per degree", is plotted against Φ (in degrees) in Fig. 7. We note the following two facts: (i) $\alpha(\Phi)$ is an increasing function of Φ; and (ii) $\alpha(\Phi)$ varies

[16] If n^* is the "star" function derived from n_1, and $J^*(n_1^*)$ the corresponding "J^*", it is not difficult to show that $J^*(n^*) < 0$. If A_1^+ is the portion of A_1 which can be seen from $-\infty$, $J^*(n_1^*) = \dfrac{1}{\pi}\{2A_1^+ - (A_1 + A_2)\}$.

remarkably little, the variation being only a quarter (approximately) of its minimum.

Now the domain W_0^* is limited to a considerable extent by the shape of the lines t_0. Either W_0^* is schlicht, in which case $J^* \leqq 0$; or W_0^* covers some region D of the w-plane twice, in which case we suppose that D lies inside the region $\Delta(\Phi)$ and touches the two boundary lines $t_0(\varphi)$, $t_0(\varphi')$ of $\Delta(\Phi)$. Then every point of $\Delta(\Phi)$ (and its reflection in the real axis) is covered at least once. If now Φ is small, then the resulting contribution to J^* is below average by (i); if Φ is large, the contribution is relatively large, but mean 1-valency prevents too many

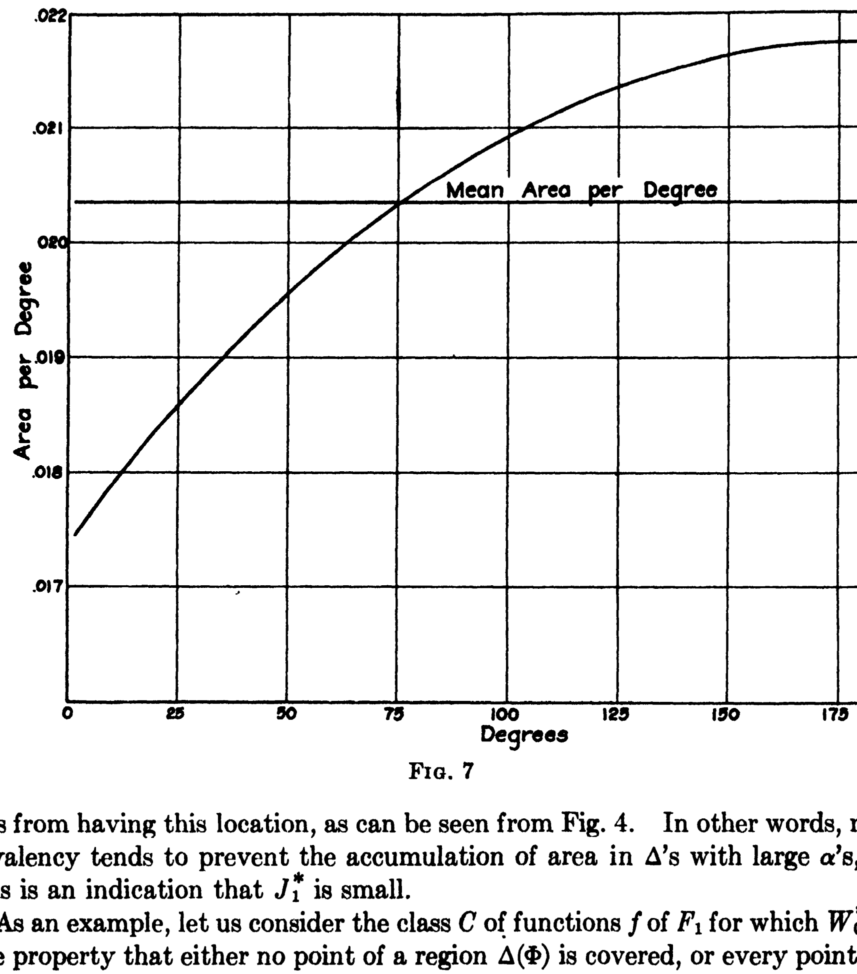


FIG. 7

D's from having this location, as can be seen from Fig. 4. In other words, mean 1-valency tends to prevent the accumulation of area in Δ's with large α's, and this is an indication that J_1^* is small.

As an example, let us consider the class C of functions f of F_1 for which W_0^* has the property that either no point of a region $\Delta(\Phi)$ is covered, or every point of it is covered twice. Then by mean 1-valency at most 180 regions $\Delta(\Phi)$ are occupied (since each must be filled twice). If we fill twice the regions $\Delta(\Phi)$, $\Phi = 90° + v$, $v = 1°, 2°, \cdots, 180°$, a precise computation shows that (10.6) is then true for this function with $A = 4.35$; and by property (i) above we see at once that (10.6) with $A = 4.35$ is true throughout the class C. For this result we have used mean

1-valency only on the circumference $|w| = 1$. If we use it everywhere, we find that (10.6) is true in C with $A = 4$.

11. In conclusion I should like to point out that results obtained for mean 1-valent functions may be immediately transferred to mean p-valent functions of the form

$$f(z) = \sum_{n=p}^{\infty} a_n z^n. \tag{11.1}$$

In fact, the function

$$f_1(z) = \{f(z)\}^{1/p}$$

is regular in $|z| < 1$ since the mean p-valency of f implies that $f = 0$ only at $z = 0$. Let $n_1(w)$ be the number of times f_1 takes the value w in $|z| < 1$, and let $p_1(R)$ be defined by (2.2) and (2.3) in terms of n_1. Then if $p(R)$ is the corresponding function for f, it is easy to see that

$$p_1(R) = \frac{1}{p} \cdot p(R^p), \tag{11.2}$$

and so

$$\begin{aligned} \int_0^R p_1(R)\, d(\pi R^2) &= \frac{1}{p} \int_0^R p(R^p)\, d(\pi R^2) \\ &= \frac{1}{p} \int_0^{R^p} p(R)\, d(\pi R^{2/p}) \leqq \frac{1}{p} \cdot p\pi R^2 = \pi R^2 \end{aligned} \tag{11.3}$$

by Lemma 1 of my paper (4). Hence f_1 is mean 1-valent, and so we have that $|a_{p+1}| \leqq 2p \cdot |a_p|$, $d = \operatorname*{Inf}_{|z|<1} |f(z)/z^p| > |a_p|/7^p$, etc.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

(1) Bermant, A., "Remarque sur le lemme de Schwarz," *C. R. Acad. Sci. Paris*, 207 (1938), 31–33.

(2) Prawitz, H., "Über Mittlewerte analytischer Funktionen," *Arkiv för Mat.*, 20A (1927–28), 6, 1–12.

(3) Spencer, D. C., "On finitely mean valent functions," *Proc. London Math. Society* (to appear shortly).

(4) Spencer, D. C., "On finitely mean valent functions II," *Trans. Amer. Math. Society*, 48 (1940), 418–435.

(5) Spencer, D. C., "On an inequality of Gransky," *Proc. Nat. Acad. Sciences*, 26 (1940), 616–621.

(6) Spencer, D. C., "On distortion in analytic transformations," *Journal Math. and Physics, Mass. Inst. of Technology*, 20 (1941), 124–126.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

NECESSARY CONDITIONS IN THE THEORY OF INTERPOLATION IN THE COMPLEX DOMAIN

J. H. CURTISS

(Received May 18, 1940)

1. Introduction

Let C be a closed limited point set of the complex z-plane, and let A_C denote the class of functions analytic and single-valued on C. Let

$$M: \left\{ \begin{array}{l} \alpha_1^{(1)} \\ \alpha_1^{(2)}, \alpha_2^{(2)} \\ \vdots \\ \alpha_1^{(n)}, \alpha_2^{(n)}, \cdots, \alpha_n^{(n)} \\ \vdots \qquad\qquad\quad \ddots \end{array} \right\}$$

be an infinite set of points of the finite plane which either lie on C or have no limit point except on C. (The points of M need not be all distinct.) A function $f(z) \in A_C$ will be analytic and single-valued at each of the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$, if n is sufficiently large, say $n > n(f)$. There exists a unique polynomial $L_n(z; f)$ of degree at most $n - 1$, $n > n(f)$, found by interpolation to the function $f(z)$ in the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$.[1]

Certain conditions are easily proved to be sufficient for the convergence of the sequence $\{L_n(z; f)\}$ to $f(z)$ on C. The necessity of such conditions seems first to have been studied by Kalmár[2] in his thesis of 1926. He proved the following theorem:

THEOREM A. *Let C be a closed Jordan region and let the points M lie on the boundary B of C. Then a necessary condition that $\lim_{n\to\infty} L_n(z; f) = f(z)$ uniformly on C for each function $f(z) \in A_C$ is that the points M be uniformly-densely distributed on B.*

The sufficiency of the condition was proved earlier by Fejér.[3] A uniformly-dense distribution of the points M on the Jordan curve B is defined as follows: Let the analytic function $w = \varphi(z)$ map the complement of C (with respect to the extended plane) onto the region $|w| > 1$ so that the points at infinity

[1] See J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, New York, 1935, pp. 49-51.

[2] L. Kalmár, Über Interpolation. (In Hungarian.) Matematikai és Physikai Lapok, 1926, pp. 120-149.

[3] L. Fejér, Interpolation und konforme Abbildung, Göttinger Nachrichten, 1918, pp. 319-331.

correspond to each other. The mapping is bi-uniform and continuous for $|w| \geqq 1$. The points M are said to be uniformly-densely distributed on B if the points $\varphi(\alpha_k^{(n)})$, $k = 1, \cdots, n$, $n = 1, 2, \cdots$, are uniformly-densely distributed on the unit circle in the sense of Weyl;[4] that is, if for any arc of length l on the unit circle,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\substack{k \\ \varphi(\alpha_k^{(n)}) \subset l}} 1 = \frac{l}{2\pi}.$$

Kalmár[5] also proved the following theorem in the case that the region K is simply connected:

THEOREM B. *Let the complement K of C be connected and regular in the sense that K possesses a Green's function $G(x, y)$ with pole at infinity. Let $\varphi(z) = e^{G+iH}$, where $H = H(x, y)$ is a harmonic conjugate of G, and let Δ be the transfinite diameter (or capacity) of C. Then a necessary condition that $\lim_{n\to\infty} L_n(z; f) = f(z)$ uniformly on C, for each function $f(z) \epsilon A_C$, is that*

$$\lim_{n\to\infty} |\omega_n(z)|^{1/n} = \Delta\, |\varphi(z)|, \qquad \omega_n(z) = \prod_{k=1}^{n} (z - \alpha_k^{(n)}),$$

uniformly for z on any closed limited point set interior to K.

The sufficiency of the condition was also proved by Kalmár in the case that K is simply connected. The formulation of Theorem B given here, and the proof of the theorem and its converse in the more general case, are due to Walsh,[6] who has also generalized Theorem A and its converse to the case of any finite number of mutually exterior closed Jordan regions.[7]

In the proof of these theorems, Kalmár and Walsh make use of the uniform convergence on C of the sequence $\{L_n(z; f)\}$ only for functions $f(z)$ belonging to a certain subclass A_C' of A_C. This subclass A_C' is the class of all functions of the type $f(z) = 1/(t - z)$, where the parameter t assumes the affix of every point in K. It is our main purpose in this paper to show that the conclusions of Theorem A and Theorem B (with certain restrictions on the set C in the latter case) can be derived from an initial assumption which is much weaker than the condition of uniform convergence of the sequence $\{L_n(z; f)\}$ for each $f(z) \epsilon A_C'$. In fact, our hypothesis will be merely that at a single interior point z_0 of C, the following inequality holds:

$$\overline{\lim_{n\to\infty}}\, |L_n(z_0; f)|^{1/n} \leqq 1, \qquad f(z) \epsilon A_C'. \tag{1.1}$$

[4] H. Weyl, Über die Gleichverteilung von Zahlen mod Eins, Math. Ann., vol. 77 (1916), pp. 313-352.

[5] Kalmár, loc. cit.

[6] Walsh, op. cit., pp. 154-155, 159-162.

[7] Walsh, op. cit., pp. 168-170. See also G. Szegö, Bemerkungen zu einer Arbeit von Herrn M. Fekete, Math. Zeit., vol. 21 (1924), pp. 203-208.

In the course of our discussion, we shall derive several auxiliary results which appear to be of some interest in themselves, and which will accordingly be developed in some detail They include several new non-trivial sufficient conditions for the convergence of the sequence $\{L_n(z; f)\}$ (Theorems 2.2, 4.1, 5.1, 5.2), an inequality which promises to be useful in the study of the fundamental polynomials $\omega_n(z) = \prod_{k=1}^{n} (z - \alpha_k^{(n)})$ (§3), and a new sufficient condition that points be uniformly-densely distributed on a Jordan curve (Theorem 5.1). In connection with the inequality, we shall digress briefly to discuss a recent paper by Lammel[8] (§4).

2. Preliminary Theorems

We shall show in this section that condition (1.1), without the restriction that z_0 be an interior point of C, implies the convergence of the sequence $\{L_n(z_0 ; f)\}$ for any $f(z) \epsilon A_C$.

THEOREM 2.1. *Let C be an arbitrary closed limited point set whose complement K is connected. Let the limit points of M all lie on C. Let z_0 be any point of C and let K' be an arbitrary closed limited point set interior to K. If*

$$\overline{\lim_{n \to \infty}} \, | L_n[z_0 ; 1/(t - z_0)] |^{1/n} \leqq 1, \qquad t \epsilon K, \tag{2.1}$$

then

$$\overline{\lim_{n \to \infty}} \left| \frac{\omega_n(z_0)}{\omega_n(t)} \right|^{1/n} \leqq \rho < 1, \qquad t \epsilon K', \tag{2.2}$$

where ρ is independent of t.

The proof is indirect. For a fixed t, if n is sufficiently large, we may write

$$L_n[z; 1/(t - z)] = \frac{1}{t - z}\left[1 - \frac{\omega_n(z)}{\omega_n(t)}\right], \qquad z \neq t. \tag{2.3}$$

Suppose now that for some $t_0 \epsilon K$,

$$\overline{\lim_{n \to \infty}} \left| \frac{\omega_n(z_0)}{\omega_n(t_0)} \right|^{1/n} = \mu_0 \geqq 1. \tag{2.4}$$

If $\mu_0 > 1$, we obtain a contradiction at once by reference to (2.1) and (2.3). But in any case we may proceed as follows:

The functions $(\log | \omega_n(t) |)/n$ form a normal family of harmonic functions in any closed limited sub-region of K, provided that n is sufficiently large, because in that case $| t - \alpha_k^{(n)} |$, $k = 1, \cdots, n$, is not only uniformly bounded, but also uniformly bounded from zero. Since $(\log | \omega_n(t) |)/n \to \infty$ as $t \to \infty$, no limit function of the family can be identically constant in K. Let the subscript n' run through an infinite subsequence of the positive integers such that if n in (2.4)

[8] E. Lammel, Über Approximation regulärer Funktionen eines komplexen Argumentes durch rationale Funktionen, Math. Zeit., vol. 46 (1940), pp. 104–116.

is replaced by n', the symbol $\overline{\lim}$ can be replaced by the symbol lim. There exists a limit function $W(t)$ defined by the equation

$$\log W(t) = \lim_{n''\to\infty} \frac{\log |\omega_{n''}(t)|}{n''}, \qquad t \in K, t \neq \infty,$$

where n'' assumes an infinite subsequence of the values assumed by n'. The Principle of the Maximum for harmonic functions, applied to $\log W(t)$, now shows that there exists a point $t_1 \in K$ such that $W(t_1) < W(t_0)$. We have

$$\lim_{n''\to\infty} \left| \frac{\omega_{n''}(z_0)}{\omega_{n''}(t_1)} \right|^{1/n''} = \lim_{n''\to\infty} \left| \frac{\omega_{n''}(z_0)\omega_{n''}(t_0)}{\omega_{n''}(t_0)\omega_{n''}(t_1)} \right|^{1/n''} = \mu_0 \frac{W(t_0)}{W(t_1)} > 1.$$

Thus there exists a number $\mu_1 > 1$ such that for $n'' > n(\mu_1)$, we can write the inequality

$$\left| \frac{\omega_{n''}(z_0)}{\omega_{n''}(t_1)} \right| \geqq \mu_1^{n''}.$$

Substituting into (2.3), we find that

$$|L_{n''}[z_0; 1/(t_1 - z_0)]|^{1/n''} \geqq \left| \frac{1}{t_1 - z_0} \right|^{1/n''} (\mu_1^{n''} - 1)^{1/n''}, \qquad n'' > n(\mu_1),$$

which contradicts (2.1). Thus for each t in K the first member of (2.2) is less than unity. The fact that it is bounded from unity for $t \in K'$ now follows from the observation that if n is sufficiently large, the functions whose superior limit appears in (2.2) form a bounded equicontinuous family for $t \in K'$, so the superior limit is a continuous function of t on K'.

THEOREM 2.2. *Let C be an arbitrary closed limited point set whose complement K is connected. Let the limit points of M all lie on C. A necessary and sufficient condition that*

$$(2.5) \qquad L_n(z_0; f) - f(z_0) = 0(r^n), \qquad r < 1, \qquad z_0 \in C, \qquad f(z) \in A_C,$$

where r depends on $f(z)$, is that

$$\overline{\lim_{n\to\infty}} |L_n[z_0; 1/(t - z_0)]|^{1/n} \leqq 1, \qquad t \in K.$$

The necessity is obvious from (2.3). In the sufficiency, we use the fact that given a function $f(z) \in A_C$, there exists a point set S consisting of a finite number of regions, each bounded by a finite number of non-intersecting rectifiable curves, such that S contains C in its interior and such that $f(z)$ is analytic and single-valued on the corresponding closed set $\bar{S}$.[9] Let the set K' of Theorem 2.1 be the boundary of S. The Cauchy-Hermite formula for $L_n(z; f)$ can be set up as follows:

$$L_n(z_0; f) = \frac{1}{2\pi i} \int_{K'} \frac{f(t)}{t - z} \left[1 - \frac{\omega_n(z_0)}{\omega_n(t)} \right] dt, \qquad n > n(f),$$

and the remainder of the proof follows at once from Theorem 2.1.

[9] Walsh, op. cit., pp. 12–13.

We add three remarks concerning these theorems.

(a) It is not difficult to devise sequences M for which (2.1) obviously holds. If the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$, are assumed to be distinct, we have the familiar formula

$$L_n(z_0; f) = \sum_{k=1}^{n} f(\alpha_k^{(n)}) \frac{\omega_n(z_0)}{(z_0 - \alpha_k^{(n)})\omega_n'(\alpha_k^{(n)})}.$$

Thus a sufficient condition that the sequence M should satisfy (2.1) is that

$$\left| \frac{\omega_n(z_0)}{(z_0 - \alpha_k^{(n)})\omega_n'(\alpha_k^{(n)})} \right| \leqq \chi_n, \qquad k = 1, \cdots n, \qquad \overline{\lim_{n\to\infty}}\, \chi_n^{1/n} \leqq 1. \tag{2.6}$$

Sequences M which satisfy (2.6) for every point $z_0 \in C$ are easily shown to exist, and have been studied by a number of authors; an example of such a sequence M is given by the "Fekete points."[10]

(b) If instead of being valid merely at a single point z_0, (2.1) or (2.2) holds at each point z of a closed subset C' of C containing no limit points of M, then the functions whose superior limit appears in (2.2) form (for n sufficiently large) a bounded equicontinuous family in the two variables z and t, $z \in C'$, $t \in K'$. Accordingly, in this case the number ρ in (2.2) may be chosen so as to be independent of both z and t, $z \in C'$, $t \in K'$, and (2.5) holds uniformly for $z \in C'$.

(c) The proof of Theorem 2.1 does not apply to the case in which K is not connected, because then it may happen that the limit function $W(t)$ used in the proof is a constant. In fact, both Theorems 2.1 and 2.2 are false in this case, as may be seen by letting C be a circular annulus whose outer boundary is the unit circle, and letting the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$ be the n-th roots of unity.

3. An inequality

LEMMA 3.1. *Let $a_1, a_2, \cdots, a_n$ be n real numbers such that $q\eta \leqq a_k \leqq 1$, $k = 1, \cdots, n$, where $0 < \eta < 1$, $0 < q < 1/\eta$. Then*

$$q(a_1, a_2, \cdots, a_n) = \prod_{k=1}^{n} (a_k + \eta)^{1/n} - \prod_{k=1}^{n} a_k^{1/n}$$

$$\leqq \begin{cases} \eta, & \eta^2 + \eta \geqq 1/(q+1) \\ (q\eta)^{\kappa} \dfrac{\log q - \log (q+1) - \log (\eta + 1)}{\log \eta + \log (q+1) + \log (\eta+1)}, & \eta^2 + \eta < 1/(q+1), \end{cases} \tag{3.1}$$

where

$$\kappa = \frac{\log \log q\eta - \log [\log \eta + \log (q+1) + \log (\eta+1)] - \log (\eta+1)}{\log (q+1) - \log q - \log (\eta+1)}.$$

It should be observed that the right member of the inequality is independent of n.

[10] See Walsh, op. cit., pp. 170–173.

To prove the inequality, we first remark that the maximum value of the function $g(a_1, a_2, \cdots, a_n)$ in the n-dimensional cube $q\eta \leqq a_k \leqq 1, k = 1, \cdots, n$, is assumed at one of the "corners" of this cube. This may be readily established by using the symmetry of g and by examining the partial derivatives of the function $g(x_1, x_2, \cdots, x_k, a_{k+1}, \cdots, a_n)$ with respect to the a's, where each of the x's assumes either the value $q\eta$ or 1. We leave the details to the reader.

Suppose now that each a is an x, and that m of the x's are equal to $q\eta$ and $n - m$ of them are equal to 1. Then

$$g(x_1, x_2, \cdots, x_n) = h(m) = (1 + \eta)^{(n-m)/n}(q\eta + \eta)^{m/n} - (q\eta)^{m/n}.$$

Let the range of variation of m be extended to include all the real numbers. Computing the derivative of $h(m)$ by the methods of elementary calculus, we find at once that if $\log(1 + \eta) + \log(q\eta + \eta) \geqq 0$, then $h'(m) > 0$ for all m, so in this case the maximum of $h(m)$ in the interval $0 \leqq m \leqq n$ occurs when $n = m$. Since $h(n) = \eta$, we obtain the first part of (3.1). On the other hand, if $\log(1 + \eta) + \log(q\eta + \eta) < 0$, then there exists a unique root $m = \hat{m}$ of the equation $h'(m) = 0$ in the interval $0 < m < n$, and $\hat{m}$ is readily found to be the maximum point of $h(m)$. The rather artificial-looking expression in the second part of (3.1) is simply $h(\hat{m})$, and $\kappa = \hat{m}/n$.

It is apparent from the method of proof and from the order of magnitude of $h(m) - h([m])$ (where $[m]$ denotes the largest integer in m) that the inequality cannot be essentially improved. The inequality is more convenient when stated in less explicit form. Since κ is positive, the second expression in the third member of (3.1) is clearly dominated by $[\log 2 + \log(1 + 1/q)]/|\log \eta|$. Thus we obtain the following corollary:

LEMMA 3.2. *Let $a_1, \cdots, a_n$, η, and q satisfy the conditions in Lemma 3.1. Then there exists a number A_q independent of n and η such that*

$$\prod_{k=1}^{n} (a_k + \eta)^{1/n} - \prod_{k=1}^{n} a_k^{1/n} \leqq \frac{A_q}{|\log \eta|}. \tag{3.2}$$

It happens that the right member of (3.2) reflects accurately the order of magnitude of the right member of (3.1) for small values of η, because $\lim_{\eta \to 0} (q\eta)^\kappa = 1/e$.

We now apply these results to the fundamental polynomials $\omega_n(z)$.

THEOREM 3.1. *Let z be a point of the complex z-plane such that $z \neq \alpha_k^{(n)}$, $k = 1, \cdots, n$. Let D be a number such that $|z - \alpha_k^{(n)}| \leqq D$, $k = 1, \cdots, n$. If t is a point which satisfies the inequality $|t - z| \leqq \delta$, where $0 < \delta < D$, and if q be chosen so that $q\delta \leqq |z - \alpha_k^{(n)}|$, $k = 1, \cdots, n$, then a number $A_{q,D}$ exists which is independent of n, z, t, and δ, such that*

$$|\omega_n(t)|^{1/n} \leqq |\omega_n(z)|^{1/n} + \frac{A_{q,D}}{|\log \delta|}.$$

For the proof, let $a_k = |z - \alpha_k^{(n)}|/D$, $k = 1, \cdots, n$, and $\eta = \delta/D$. The numbers a_k, η, and q satisfy the conditions of Lemma 3.1. Since $|t - \alpha| \leqq |z - \alpha| + \delta$, we may write

$$|\omega_n(t)|^{1/n} = \prod_{k=1}^{n} |t - \alpha_k^{(n)}|^{1/n} \leqq \prod_{k=1}^{n} (|z - \alpha_k^{(n)}| + \delta)^{1/n}$$

$$= |\omega_n(z)|^{1/n} + D\left[\prod_{k=1}^{n} (a_k + \eta)^{1/n} - \prod_{k=1}^{n} a_k^{1/n}\right].$$

The conclusion of the theorem now follows at once from Lemma 3.2.

THEOREM 3.2. *Let C be an arbitrary closed point set of the extended plane (not the entire plane) with interior points, and let the derivative of the set M be a limited set contained in the boundary of C. Let $\{z_j\}$ be an infinite sequence of finite interior points of C approaching a finite boundary point β of C. Then given any $\epsilon > 0$, there exists a corresponding sequence $\{t_j\}$ of points of the complement of C approaching β, such that for all $j > j(\epsilon)$,*

$$|\omega_n(t_j)|^{1/n} \leqq |\omega_n(z_j)|^{1/n} + \epsilon, \qquad n > n(j). \tag{3.3}$$

For let $D = \overline{\text{bound}}\, |z_j - \alpha_k^{(n)}|, j = 1, 2, \cdots, k = 1, 2, \cdots, n, n = 1, 2, \cdots$. Let $3\epsilon_j/4$ denote the distance from z_j to the boundary of C. By choosing j sufficiently large, say $j > j(\epsilon)$, we can make ϵ_j so small that (in the notation of Theorem 3.1) $A_{1/2,D}/|\log \epsilon_j| < \epsilon$. There exists for each j a point t_j in the complement of C such that $|t_j - z_j| < \epsilon_j$. Moreover, for n sufficiently large, say $n > n(j)$, $|z_j - \alpha_k^{(n)}| \geqq \frac{1}{2}\epsilon_j$, $k = 1, \cdots, n$. The conditions of Theorem 3.1 with $q = \frac{1}{2}$, $\delta = \epsilon_j$, are satisfied for each $j > j(\epsilon)$, provided that $n > n(j)$, and accordingly we obtain (3.3).

In general, t_j and z_j cannot be interchanged in (3.3); that is, given a sequence $\{z_j\}$ in C approaching a boundary point of C, we cannot conclude that there exists a sequence $\{t_j\}$ in the complement of C such that for $j > j(\epsilon)$,

$$\left||\omega_n(z_j)|^{1/n} - |\omega_n(t_j)|^{1/n}\right| \leqq \epsilon, \qquad n > n(j), \tag{3.4}$$

where $\epsilon > 0$ is arbitrary. That this situation obtains can easily be proved by examples; it is attributable to the fact that an inequality such as (3.1), where the right member approaches zero with η, is not available if $q = 0$. However, if C is a closed Jordan region bounded by a suitably restricted curve (the requirements fall considerably short of analyticity), then the equicontinuity property of the sequence $\{|\omega_n(z)|^{1/n}\}$ expressed by (3.4) can be established by means of Theorem 3.1. We shall not need this equicontinuity property in the investigation of this sequence which follows, so we shall not discuss the matter further.

4. Applications of the inequality

Let C be a closed limited point set whose complement K is connected; furthermore, let us assume that C has interior points and that all limit points of the

set M lie on the boundary B of C. Suppose now that (1.1) holds at an interior point z_0 of C. There exists a maximal region R_0 of C containing z_0; the boundary of R_0 is contained in B. If n is sufficiently large, the functions $(\log |\omega_n(z)|)/n$ form a normal family of harmonic functions in any closed subregion of R_0. It is a simple consequence of Theorem 2.1 and 3.2 that each limit function of this family, or of the family $|\omega_n(z)|^{1/n}$, is a constant, as we now shall show.

Let $W(z)$ be an arbitrary limit function of the family $|\omega_n(z)|^{1/n}$; $W(z)$ is defined by an equation of the following type:

$$\log W(z) = \lim_{n' \to \infty} \frac{\log |\omega_{n'}(z)|}{n'}, \qquad z \in R_0,$$

where n' runs through some infinite subsequence of the positive integers. If $W(z)$ is not identically constant in R_0, then by the Principle of the Maximum for harmonic functions, there is an infinite sequence $\{z_j\}$ of points of R_0 approaching a point of B such that for some $\epsilon > 0$, $W(z_j) < W(z_0) - 2\epsilon$, $j = 1, 2, \cdots$. But by Theorem 3.2, there exists a subscript J and a point t_J in K such that

$$|\omega_n(t_J)|^{1/n} \leqq |\omega_n(z_J)|^{1/n} + \epsilon, \qquad n > n(J).$$

Therefore

$$\overline{\lim_{n' \to \infty}} |\omega_{n'}(t_J)|^{1/n'} \leqq \lim_{n' \to \infty} |\omega_{n'}(z_J)|^{1/n'} + \epsilon = W(z_J) + \epsilon < W(z_0) - \epsilon;$$

$$\overline{\lim_{n' \to \infty}} |\omega_{n'}(t_J)|^{1/n'} < \lim_{n' \to \infty} |\omega_{n'}(z_0)|^{1/n'}.$$

But the last inequality contradicts (2.2), so $W(z)$ must be a constant. We are now in a position to state the following theorem:

THEOREM 4.1. *Let C be a closed limited point set whose complement K is connected. Let all limit points of the set M lie on the boundary of C. Let z_0 be some interior point of C, and let R_0 be the maximal region of C containing z_0. If*

$$\overline{\lim_{n \to \infty}} |L_n[z_0; 1/(t - z_0)]|^{1/n} \leqq 1, \qquad t \in K,$$

then

(a) *each limit function in R_0 of the normal family* $(\log |\omega_n(z)|)/n$ *is a constant;*

(b) *for every function $f(z) \in A_C$, $L_n(z; f) - f(z) = 0(r^n)$, $r < 1$, uniformly on any closed point set interior to R_0.*

Part (b) of the theorem is a simple consequence of part (a) of the theorem, and of Theorems 2.1, 2.2, and Remark (b) of §1. We leave the proof to the reader.

The theorem is false if limit points of M are allowed to lie inside R_0. For

example, if C is the set $|z| \leqq 1$, and if $\alpha_k^{(2n)} = e^{2\pi ik/n}$, $k = 1, \cdots, n$, $\alpha_k^{(2n)} = \frac{1}{2}$, $k = n + 1, \cdots, 2n$, then

$$\lim_{n\to\infty} |\omega_{2n}(z)|^{1/2n} = \begin{cases} |z - \frac{1}{2}|^{\frac{1}{2}}, \ |z| < 1 \\ (|z|\,|z - \frac{1}{2}|)^{\frac{1}{2}}, \ |z| > 1; \end{cases}$$

$$\lim_{n\to\infty} L_{2n}(z; f) = f(z), \ |z - \tfrac{1}{2}| \leqq \tfrac{1}{2}, f(z) \in A_C;$$

but the sequence $\{L_{2n}(z; f)\}$, $f(z) \in A_C$, may diverge if $|z - \frac{1}{2}| > \frac{1}{2}$.

The problem involved in the proof of part (a) of the theorem is typical of a class of problems arising in the study of interpolation by rational functions in the complex domain, in which certain properties of the sequence $\{|\omega_n(z)|^{1/n}\}$ are to be derived from other properties of this sequence known for z both interior and exterior to a closed set C. In dealing with such problems, the inequality (3.3) may provide a useful tool.[11] The writer intends to discuss this matter elsewhere in more detail. However, it seems not out of place to mention an application of the inequality to a line of investigation which is related to Theorem A, but which is concerned with interpolatory series of rational functions with poles at the points M.

Recently Lammel[12] has derived necessary and sufficient conditions that a certain expansion of the type just described should converge interior to a Jordan region R to any function $f(z) \in A_{\bar{R}}$, where $\bar{R}$ is the corresponding closed region. The points M are chosen on the boundary of R. In studying the necessity, his starting point essentially is the condition

$$\overline{\lim_{n\to\infty}} \frac{|z - a|}{|\omega_n(z)|^{1/n}} \leqq \underline{\lim_{n\to\infty}} \frac{|t - a|}{|\omega_n(t)|^{1/n}}, \tag{4.1}$$

which is to hold for all $z \in R$ and all t in the complement K of $\bar{R}$, where a is some point of R. From this inequality he infers, by a rather complicated argument, that

$$\lim_{n\to\infty} \frac{|t - a|}{|\omega_n(t)|^{1/n}} = 1, \tag{4.2}$$

uniformly on any closed subset of K. But a line of reasoning similar to that used to establish Theorem 4.1a (the reader will have no trouble in supplying the details) will show that (4.2) is quite an immediate consequence of (4.1) and Theorem 3.2. Moreover in deriving (4.2) by this method, the restriction to a Jordan region becomes quite superfluous; indeed, R may be any region whose boundary is also the boundary of an infinite region.[13]

---

[11] Another application of the inequality similar to the one in this section appears in the proof of Theorem 5.1b below.

[12] E. Lammel, loc. cit.

[13] It follows from this that Lammel's Theorem 1 (loc. cit., p. 105) is true for such regions R.

5. The revised version of Theorems A and B

We shall now prove Theorem A under the revised hypothesis (1.1), and also Theorem B under this hypothesis, but with the necessary additional restriction on C that it be the closure of a region R whose boundary is also the boundary of an infinite region. Our method of approach is indicated in the following theorem:

THEOREM 5.1. *Let R be a limited region whose boundary B is the boundary of an infinite region K. Let all limit points of the set M lie on B. Necessary conditions that every limit function in R of the normal family* $(\log |\omega_n(z)|)/n$ *should be a constant are that*

(5.1) (a) $\lim_{n\to\infty} |\omega_n(z)|^{1/n} = \Delta$, *uniformly on any closed point set interior to R, where Δ is the transfinite diameter of $R + B$,*

(5.2) (b) $\lim_{n\to\infty} |\omega_n(z)|^{1/n} = \Delta |\varphi(z)|$, *uniformly on any closed limited point set interior to K, where $w = \varphi(z)$ maps K conformally onto the region $|w| > 1$ so that the points at infinity correspond to each other.*

If R is a Jordan region and if all the points M lie on B, then a further necessary condition is that the points M are uniformly-densely distributed on B.

Put in another way, the theorem states that if the mean radii of the lemniscates $|\omega_n(z)| = \mu_n$, $n = 1, 2, \cdots$, passing through an arbitrary point $z \in R$, approaches as $n \to \infty$ a limit which is constant for all $z \in R$, then this limit must be the same as the limit of the mean radii of the minimal covering lemniscates of $R + B$ of respective degrees n.[14]

Let $\zeta_n(z) = (\log |\omega_n(z)|)/n$. Let $\{\zeta_{n'}(z)\}$ be an arbitrary convergent subsequence of the family $\zeta_n(z)$. The convergence of this subsequence is uniform in some neighborhood N of a point $z_1 \in R$, and the neighborhood is independent of the particular subsequence under consideration. Since the limit of the subsequence is a constant, the two corresponding subsequences of partial derivatives with respect to x and y, $z = x + iy$, both converge uniformly to zero in N.[15] Therefore

$$\lim_{n'\to\infty} \frac{\sum_1^{n'} (z - \alpha_k^{n'})^{-1}}{n'} = 0, \quad \text{uniformly}, \quad z \in N. \tag{5.3}$$

But the sequence $\{\zeta_{n'}(z)\}$ was arbitrary, so we are able to conclude that every limit function of the normal family of analytic functions $[\sum_1^n (z - \alpha_k^{(n)})^{-1}]/n$ is identically zero in N. Therefore (5.3) is still true if n' is replaced by n.

[14] The latter limit is known to be Δ. See M. Fekete, Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen, etc., Math. Zeit., vol. 17 (1923), pp. 228-249.

[15] W. F. Osgood, Lehrbuch der Funktionentheorie, fifth edition, Leipzig, 1928, vol. I, p. 682.

Differentiating and using the Weierstrass Double Series Theorem, we obtain the equations

$$(5.4)\qquad \lim_{n\to\infty} \mathcal{R}\left[\frac{A_\nu \sum_{k=1}^{n} (z_1 - \alpha_k^{(n)})^{-\nu}}{n}\right] = 0, \qquad \nu = 1, 2, 3, \cdots,$$

where the numbers A_ν are any complex constants.

Let $z = \psi(w)$ be the inverse of the mapping function $w = \varphi(z)$. The function $\psi(w)$ is analytic and single-valued for $|w| > 1$ and may be represented by a Laurent series of the following type:

$$z = \psi(w) = cw + \sum_{J}^{\infty} C_k w^{-k}, \qquad c \neq 0, \qquad |w| > 1,$$

where $|c| = \Delta$.[16] Thus

$$(5.5)\qquad \lim_{z\to\infty}\left|\frac{z}{\varphi(z)}\right| = \left|\frac{1}{\varphi'(\infty)}\right| = \Delta.$$

Therefore the function $l(z) = \log|(z_1 - z)/\varphi(z)|$, when assigned the value $\log \Delta$ at infinity, is harmonic for $z \in K$; and if $|\varphi(z)|$ is defined to be unity on B, $l(z)$ is continuous in $K + B$. Invert the plane in the point z_1 by means of the transformation $z' = 1/(z_1 - z)$, and let K' be the image of K. The function $l_1(z') = l(z_1 - 1/z')$ is harmonic interior to the limited region K' and continuous in the corresponding closed region $\bar{K}'$. A theorem of Walsh[17] now states that given an arbitrary $\epsilon > 0$, there exists a harmonic polynomial $P(z')$ such that $|l_1(z') - P(z')| < \epsilon/4$, $z' \in \bar{K}'$. The polynomial $P(z')$ is an expression of the type $A_0 + \sum_1^m \mathcal{R}(A_\nu z'^\nu)$, where A_0 is a real constant and the A_ν are complex constants. In particular, we find that $|l_1(0) - P(0)| = |\log \Delta - A_0| < \epsilon/4$. If we use this fact when we reverse the inversion, we obtain the inequality

$$\left|l(z) - \log \Delta - \sum_{\nu=1}^{m} \mathcal{R}[A_\nu (z_1 - z)^{-\nu}]\right| < \frac{\epsilon}{2}, \qquad z \in K + B,$$

or

$$\left|\log|z_1 - z| - \log \Delta - \sum_{\nu=1}^{m} \mathcal{R}[A_\nu (z_1 - z)^{-\nu}]\right| < \frac{\epsilon}{2}, \qquad z \in B.$$

Therefore

$$\left|\frac{1}{n}\sum_{k=1}^{n} \log|z_1 - \alpha_k^{(n)}| - \log \Delta - \frac{1}{n}\sum_{\nu=1}^{m} \mathcal{R}\left[A_\nu \sum_{k=1}^{n} (z_1 - \alpha_k^{(n)})^{-\nu}\right]\right| < \frac{\epsilon}{2},$$
$$n = 1, 2, \cdots.$$

---

[16] Fekete, loc. cit., pp. 237–240.

[17] J. L. Walsh, The approximation of harmonic functions by harmonic polynomials and by harmonic rational functions, Bull. Am. Math. Soc., vol. 35 (1929), pp. 499–544; p. 503.

Referring now to (5.4), we find that for all values of n sufficiently large, $|(\log|\omega_n(z_1)|)/n - \log \Delta| < \epsilon$. Thus all the constant limit functions of the family $(\log|\omega_n(z)|/n$ have the value $\log \Delta$, which means that $\lim_{n\to\infty}(\log|\omega_n(z)|)/n = \log \Delta$, $z \in R$. Since the functions form a normal family in R, this limit is uniform on any closed point set interior to R. We have established part (a) of the theorem.

A suitably chosen branch of the function $[\omega_n(z)]^{1/n}/(\Delta\cdot\varphi(z))$, when defined to have the value unity at infinity (in accordance with (5.5)), is analytic in any closed subregion of K if n is sufficiently large, and is uniformly bounded in modulus in K, the bound being independent of n. Thus the functions $[\omega_n(z)]^{1/n}/(\Delta\varphi(z))$ form a normal family of analytic functions in any closed subregion of K, if n is sufficiently large. Let $\theta(z)$ be a limit function of this family. Since $|\theta(\infty)| = 1$, it follows from the Principle of the Maximum that either $|\theta(z)| \equiv 1$, or else there exists an infinite sequence $\{t_j\}$ of points of K approaching a point of B such that $\lim_{j\to\infty}|\theta(t_j)| < 1$. But Theorem 3.2 and part (a) of the present theorem show that the latter case is impossible. For if such a sequence $\{t_j\}$ existed, then (since $\lim_{j\to\infty}\Delta|\varphi(t_j)||\theta(t_j)| < \Delta$) there would exist a number $\epsilon > 0$ such that $|\omega_{n'}(t_j)|^{1/n'} < \Delta - 2\epsilon$, $n' \geqq n'(j)$, $j > j(\epsilon)$, where n' runs through some infinite subsequence of the positive integers. However, by Theorem 3.2, there exists a subscript $J > j(\epsilon)$ and a point $z_J \in R$ such that

$$|\omega_{n'}(z_J)|^{1/n'} \leqq |\omega_{n'}(t_J)|^{1/n'} + \epsilon < \Delta - 2\epsilon + \epsilon, \qquad n' > n'(J).$$

If we apply part (a) of the present theorem to the first member of this inequality, we obtain a contradiction at once. Thus every limit function of the normal family $[\omega_n(z)]^{1/n}/(\Delta\varphi(z))$ is a constant of modulus unity. This establishes part b of the theorem.

The proof of the last part of the theorem is now a consequence of the work of Kalmár,[18] who showed in his proof of Theorems A and B that if R is a Jordan region and if the points M lie on B, then equation (5.2) implies that the points M are uniformly-densely distributed on B.

Combining Theorems 4.1 and 5.1, we may state our revised version of Theorems A and B very briefly as follows:

THEOREM 5.2. *Under the hypotheses of Theorem* 5.1 *concerning R and M, the conditions* (a) *and* (b) *of that theorem and the uniformly dense distribution of the points M are necessary conditions that at some point $z_0 \in R$,*

$$\varlimsup_{n\to\infty} |L_n[z_0; 1/(t - z_0)]|^{1/n} \leqq 1, \qquad t \in K. \tag{5.6}$$

Of course these necessary conditions are also sufficient that (5.6) hold. In-

[18] Kalmár, loc. cit. See also Walsh, op. cit., pp. 168–170.

deed, in the light of Theorem B, it is seen that (5.6) or the hypotheses of Theorem 5.1 are sufficient conditions that

$$\lim_{n\to\infty} L_n(z;f) = f(z), \qquad z \in R + B, \qquad f(z) \in A_{R+B},$$

the convergence being not only uniform with respect to z, but "maximal."[19]

The conclusions of Theorem 5.2 would be false if z_0 were a boundary point of R, or if $R + B$ were replaced by any more general closed set of C of which z_0 is an interior point, even if the complement of C were simply connected. The reader will have no trouble verifying these statements by means of trivial examples similar to the one in §4.

It is perhaps of interest to observe that an alternative method of proof for Theorem 5.1 is available which in no way depends on Kalmár's work. The procedure consists in using a well-known necessary and sufficient condition of Weyl[20] that the points $\varphi(\alpha_k^{(n)})$ be uniformly-densely distributed on the unit circle, together with equations (5.4), to establish the last part of the theorem first. Equations (5.1) and (5.2) for a Jordan region may then be obtained readily from the Gauss Mean Value Theorem. The validity of these equations can then be extended to the more general case by approximating the region R by means of its Green's level curves.

Cornell University.

[19] Walsh, op. cit., p. 154.
[20] Weyl, loc. cit., pp. 313–316.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

HILBERT DISTANCES AND POSITIVE DEFINITE FUNCTIONS

BY S. BOCHNER

(Received January 20, 1941)

I. HILBERT DISTANCES

If $\mathfrak{S}$ is any topological space then a *continuous* complex function of two variables $f(P, Q)$ on $\mathfrak{S} \times \mathfrak{S}$ shall belong to class $\mathfrak{P}$ (positive definite functions) if

1) $f(P, Q) = \overline{f(Q, P)}$
2) $f(P, P) = f(Q, Q)$

and

3′) for any integer n, any points $P_1, \cdots, P_n$ and any complex numbers $\rho_1, \cdots, \rho_n$ the relation

$$\sum_{i,j=1}^{n} f(P_i, P_j)\rho_i\bar{\rho}_j \geqq 0, \tag{1}$$

holds.

The wider class $\mathfrak{P}_0$ shall consist of all those functions for which, other assumptions being unchanged, the decisive inequality (1) is assumed to hold only subject to the restriction

$$\sum_{i=1}^{n} \rho_i = 0. \tag{2}$$

Just how much wider the class $\mathfrak{P}_0$ actually is will be established under further assumptions. The *real* classes $\mathfrak{P}$ and $\mathfrak{P}_0$ shall consist of all real-valued functions of these classes.

We will call a function $\rho(P, Q)$ a distance function if $\rho(P, Q) = \rho(Q, P) \geqq 0$, $\rho(P, P) = 0$ and $\rho(P, Q) + \rho(Q, R) \geqq \rho(P, R)$. We call $\rho(P, Q)$ a *proper* distance if $P \neq Q$ implies $\rho(P, Q) > 0$. Introducing the real Hilbert space $\mathfrak{H}$ of sequences $\{x_n\}$, $\{y_n\}$ with the customary distance $\left(\sum_n (x_n - y_n)^2\right)^{1/2}$, we now call $\rho(P, Q)$ a *Hilbert distance* on $\mathfrak{S}$ if it is possible to map $\mathfrak{S}$ into $\mathfrak{H}$ in such a way that the value of $\rho(P, Q)$ shall be equal to the value of the latter distance for the transforms of P and Q. The following decisive criterion has been established by K. Menger and I. J. Schoenberg:[1] *if $\mathfrak{S}$ is separable, then $\rho(P, Q)$ is a Hilbert distance if and only if* $-\rho(P, Q)^2$ belongs to $\mathfrak{P}_0$. Also, Schoenberg discovered the fact that a function $f(P, Q)$ belongs to $\mathfrak{P}_0$ if and only if $e^{\lambda f(P,Q)}$ belongs to $\mathfrak{P}$ for each

[1] See I. J. Schoenberg, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. 44 (1938), 522–536.

[2] Positive definite functions on non-compact but commutative groups have been analyzed recently in A. Powzner, Doklady U.S.S.R., 28 (1940) 294–295, and D. Raikow, Doklady U.S.S.R., 27 (1940), 324–327 and 28 (1940), 296–300.

positive λ, and he analyzed Hilbert distances on open euclidean and Hilbert spaces.

In the present note we will investigate Hilbert distances for some types of *general* but *compact* separable spaces. We will obtain new relations between the classes $\mathfrak{P}$ and $\mathfrak{P}_0$ and structural criteria in terms of expansions in orthogonal systems.[2]

II. Compact Spaces with Bounded Measure

Since $f(P, Q)$ is continuous, if $\mathfrak{S}$ is a finite interval, say, then condition 3′) for $\mathfrak{P}$ can be replaced by the following condition:

3) for each integrable $\rho(P)$ the relation

$$\int\int f(P, Q)\rho(P)\rho(Q)\, dP\, dQ \geqq 0 \tag{3}$$

holds;[3] and for $\mathfrak{P}_0$ relation (3) holds subject to the restriction

$$\int \rho(P)\, dP = 0. \tag{4}$$

This replacement of sums by integrals is admissible for more general spaces.

Lemma 1. *Conditions* 3) *and* 3′) *are equivalent if* α), $\mathfrak{S}$ *is both separable and compact,* β) *there exists a Lebesque measure dP on* $\mathfrak{S}$ *for which* γ) *all Borel sets are measurable,* δ) *every open set has non-vanishing measure and* ϵ) *the total space has measure* 1.

Proof: We first observe that $\mathfrak{S}$ and $\mathfrak{S} \times \mathfrak{S}$ are each bi-compact, and that therefore every continuous function $f(P, Q)$ is uniformly continuous and bounded. In the proof M will be independent of ϵ. We will give the proof of the lemma for functions of $\mathfrak{P}_0$, for functions of $\mathfrak{P}$ it is even simpler. Let $f(P, Q)$ satisfy 3′) and let $\rho(P)$ be any function satisfying (4). Given $\epsilon > 0$ we partition $\mathfrak{S}$ into a finite number of Borel sets $\mathfrak{S}_1, \cdots, \mathfrak{S}_n$ such that $f(P, Q)$ oscillates by less than ϵ on each $\mathfrak{S}_i \times \mathfrak{S}_j$. Choosing an arbitrary point P_i in $\mathfrak{S}_i$ and putting $\rho_i = \int_{\mathfrak{S}_i} \rho(P)\, dP$ we obviously have (2) and

$$\left| \sum_{i,j=1}^{n} f(P_i, P_j)\rho_i\bar{\rho}_j - \int\int f(P, Q)\rho(P)\overline{\rho(Q)}\, dP\, dQ \right| < \epsilon M. \tag{5}$$

Letting $\epsilon \to 0$ we obtain (3). Conversely, if $f(P, Q)$ satisfies 3), and distinct points $\{P_i\}$ and numbers $\{\rho_i\}$ are given, and if (2) holds, then for $\epsilon > 0$, we pick a set of disjoint neighborhoods $\mathfrak{S}_i$ of P_i, such that $f(P, Q)$ oscillates by less than ϵ on each $\mathfrak{S}_i \times \mathfrak{S}_j$. Putting $\rho(P) = \dfrac{\rho_i}{\text{measure } \mathfrak{S}_i}$ for $P \in \mathfrak{S}_i$, and $\rho(P) = 0$ for $P \in \mathfrak{S} - (\mathfrak{S}_1 + \cdots + \mathfrak{S}_n)$, we have relation (2) and (5), and, by a limit, relation (1).

[3] Whenever the range of integration is not indicated it is the total set to which the variable refers.

In what follows we will always assume that $\mathfrak{S}$ satisfies the assumptions of the lemma.

THEOREM 1. *If $f(P, Q)$ belongs to $\mathfrak{P}$, if $\varphi(P)$ is continuous and c is a real constant, then*

$$g(P, Q) = f(P, Q) + \varphi(P) + \overline{\varphi(Q)} - c \tag{6}$$

belongs to $\mathfrak{P}_0$ except for property 2).

Conversely, if $g(P, Q)$ belongs to $\mathfrak{P}_0$, then there exists a continuous function $\varphi(P)$ and a real constant c such that

$$f(P, Q) = g(P, Q) - \varphi(P) - \overline{\varphi(Q)} + c \tag{7}$$

belongs to $\mathfrak{P}$, except for property 2), as for instance

$$\varphi(P) = \int g(P, Q)\, dQ \tag{8}$$

$$c = \int \varphi(P)\, dP = \iint g(P, Q)\, dP\, dQ. \tag{9}$$

PROOF: The first part follows from the fact that (4) implies

$$\iint \varphi(P)\rho(P)\overline{\rho(Q)}\, dP\, dQ = \int \varphi(P)\rho(P)\, dP \cdot \int \overline{\rho(Q)}\, dQ = 0; \tag{10}$$

the second part follows from the fact that (7), (8), (9) and $\lambda = \int \rho(P)\, dP$ implies

$$\iint f(P, Q)\rho(P)\overline{\rho(Q)}\, dP\, dQ = \iint g(P, Q)(\rho(P) - \lambda)\overline{(\rho(Q) - \lambda)}\, dP\, dQ.$$

We call a function $f(P, Q)$ measure invariant if $\int f(P, Q)\, dQ$ is independent of P. This leads to

THEOREM 2. *A function $g(P, Q)$ belongs to the measure invariant class $\mathfrak{P}_0$, if and only if it can be written in the form*

$$f(P, Q) - f(\Phi, \Phi) \tag{11}$$

where $f(P, Q)$ belongs to the measure invariant class $\mathfrak{P}$ and Φ is any point of $\mathfrak{S}$.

We now assume the existence of a fixed transitive group $\mathfrak{C}$ of continuous transformations of $\mathfrak{S}$ into itself, and we assume that our measure is invariant under the operations of $\mathfrak{C}$. We call a function $f(P, Q)$ group invariant if for all elements s of $\mathfrak{C}$, $f(sP, sQ) = f(P, Q)$, where sP is the transform of the point P under the transformation s. In particular $f(sP, sP) = f(P, P)$. Now if P and Q are any two points, since our group is transitive there exists an element s such that $Q = sP$. Thus from now onward property 2) of $\mathfrak{P}$ will be automatically fulfilled for any function $f(P, Q)$ which is group invariant. In order to

emphasize the independence of $f(P, P)$ of the special point P we will designate it by $f(\Phi, \Phi)$ where Φ is any fixed point which is chosen appropriately. For instance if $\mathfrak{S}$ itself is a group it is customary to take for Φ the identity element of the group.

If $f(P, Q)$ is group invariant we also have

$$\varphi(P) = \int f(P, Q)\, dQ = \int f(sP, sQ)\, dQ = \int f(sP, sQ)\, d(sQ)$$

$$= \int f(sP, Q)\, dQ = \varphi(sP),$$

and this implies that $\varphi(P)$ is a constant. In other words, if $f(P, Q)$ is group invariant it is also measure invariant. This leads to

THEOREM 3. *A function $g(P, Q)$ belongs to the group invariant class $\mathfrak{P}_0$ if and only if it can be written in the form* (11) *where $f(P, Q)$ is group invariant and belongs to $\mathfrak{P}$.*

A function $\rho(P, Q)$ is a group invariant Hilbert distance if and only if it can be represented in the form

$$\sqrt{f(\Phi, \Phi) - f(P, Q)},$$

where $f(P, Q)$ is group invariant and belongs to the real class $\mathfrak{P}$. The function $\rho(P, Q)$ is a proper distance if and only if $f(P, Q) < f(\Phi, \Phi)$ for $P \neq Q$.

III. GROUP INVARIANT FUNCTIONS

In the present section we will supplement Theorem 3 by statements involving expansions in orthonormal systems. An orthonormal system $\{\varphi_n(P)\}$ is of course defined by the property

$$\int \varphi_m(P)\overline{\varphi_n(P)}\, dP = \delta_{mn};$$

the orthonormal system we will consider will automatically consist of *continuous* functions and it will be *complete* both in the space of continuous functions and in the space of integrable functions.

The simplest space $\mathfrak{S}$ is the torus $0 \leqq x < 2\pi$ with the group of translations on it. Any continuous function $f(x, y)$ on $\mathfrak{S} \times \mathfrak{S}$ has a Fourier expansion

$$\sum_{m,n} a_{m,n} e^{imx} e^{-iny}. \tag{12}$$

Group invariance means that $f(x, y) = f(x + s, y + s)$ and this implies

$$a_{m,n} = a_{m,n} e^{i(m-n)s}.$$

Thus, $a_{m,n} = 0$ if $m \neq n$, and (12) has the form

$$\sum^{\infty} a_m e^{im(x-y)}. \tag{13}$$

Now $f(x, y)$ belongs to $\mathfrak{P}$ if and only if a_m is real and $\geqq 0$, and $\sum a_m < \infty$; $f(x, y)$ is real if and only if $a_{-m} = a_m$. Thus, by Theorem 3, $\rho(x, y)$ is a Hilbert distance if and only if it is the square root of an expression

$$(14) \qquad \sum_{m=1}^{\infty} a_m(2 - e^{im(x-y)} - e^{-i(x-y)}) = \sum_{m=1}^{\infty} a_m(e^{imx} - e^{imy})(e^{-imx} - e^{-imy})$$

where

$$(15) \qquad a_m \geqq 0 \quad \text{and} \quad \sum_m a_m < \infty.$$

If we prefer to look upon $\rho(x, y)$ as a function of the one variable $x - y$ it is more appropriate to write (15) in the form

$$(16) \qquad 2\sum_{m=1}^{\infty} a_m(1 - \cos m(x - y)) = 4\sum_{m=1}^{\infty} a_m\left(\sin \frac{m}{2}(x - y)\right)^2.$$

This result can be extended to our spaces in general in terms of expansions into "generalized spherical harmonics" as given by E. Cartan and H. Weyl.[4] If $\mathfrak{S}$ is a space satisfying lemma 1 with a transitive group $\mathfrak{G}$ of motions then there exists on $\mathfrak{S}$ a complete orthonormal system of continuous functions of the following description:

(i) corresponding to each k, $k = 0, \pm 1, \pm 2, \pm 3, \cdots$ there exist a finite rectangular system of functions $\varphi_{k,m\mu}(P)$ with

$$(17) \qquad \begin{aligned} m &= 1, \cdots, l,\ l = l(k); \\ \mu &= 1, \cdots, h,\ h = h(k), \end{aligned}$$

and a quadratic system of functions $u_{k,m\alpha}(P)$ on $\mathfrak{G}$ with $m, \alpha = 1, \cdots, l = l(k)$, such that

$$(18) \qquad \varphi_{k,m\mu}(sP) = \sum_{\alpha=1}^{l} u_{k,m\alpha}(s)\varphi_{k,\alpha\mu}(P).$$

(ii) For each k, $u_{k,m\alpha}(s)$ is an irreducible unitary representation of $\mathfrak{G}$, and for different values of k the representations are inequivalent.

(iii) for $k = 0$, $l = h = 1$, and $\varphi_{0,11}(P) = 1$, $u_{0,11}(s) = 1$; also $l(-k) = l(k)$, $h(-k) = h(k)$, and

$$\varphi_{-k,m\mu}(P) = \overline{\varphi_{k,m\mu}(P)}, \qquad u_{-k,m\alpha}(s) = \overline{u_{k,m\alpha}(s)}.$$

Now the system of functions

$$\varphi_{p,m\mu}(P) \cdot \overline{\varphi_{q,n\nu}(Q)}$$

is an orthonormal system on $\mathfrak{S} \times \mathfrak{S}$, and thus $f(P, Q)$ has an expansion

$$(19) \qquad \sum_{p,m,\mu,q,n,\nu} a_{pq,m\mu,n\nu}\,\varphi_{p,m\mu}(P)\overline{\varphi_{q,n\nu}(Q)}.$$

---

[4] H. Weyl, *Harmonics on homogeneous manifolds*, Annals of Math., 35 (1934), 486–494.

Replacing P and Q by sP and sQ respectively and using (18) we obtain as a necessary and sufficient condition for group invariance the system of relations

$$a_{pq,\alpha\mu,\beta\nu} = \sum_{m=1}^{l(p)} \sum_{n=1}^{l(q)} a_{pq,m\mu,n\nu} u_{p,m\alpha}(s) \overline{u_{q,n\beta}(s)}$$

and this is equivalent with

$$\sum_{n=1}^{l(q)} a_{pq,\alpha\mu,n\nu} u_{q,\beta n}(s) = \sum_{m=1}^{l(p)} a_{pq,m\mu,\beta\nu} u_{p,m\alpha}(s). \tag{20}$$

Since the functions $\{u(s)\}$ are linearly independent the comparison of coefficients on both sides of (20) will show that $a_{pq,m\mu,n\nu}$ vanishes if $p \neq q$ or if $m \neq n$, and that for $p = q$ and $m = n$ its value is independent of m. This leads to writing (19) in the form

$$\sum_{k=-\infty}^{\infty} \left(\sum_{\mu,\nu=1}^{h(k)} a_{k,\mu\nu} \varphi_{k,\mu\nu}(P, Q) \right) \tag{21}$$

where

$$\varphi_{k,\mu\nu}(P, Q) = \sum_{m=1}^{l(k)} \varphi_{k,m\mu}(P) \overline{\varphi_{k,m\nu}(Q)}. \tag{22}$$

We can now express the properties of $\mathfrak{P}$ in terms of the coefficients $a_{k,\mu\nu}$. Property 1) of $\mathfrak{P}$ simply means that for each k, the matrix

$$| a_{k,\mu\nu} |_{\mu,\nu=1,\cdots,h(k)} \tag{23}$$

is hermitian. Property 2) is automatically fulfilled, and property 3) means that (23) is non-negative definite. In fact putting in (3)

$$\rho(P) = \sum_{\mu=1}^{h} x_\mu \overline{\varphi_{k,m\mu}(P)}$$

we obtain

$$\sum_{\mu,\nu=1}^{h} a_{k,\mu\nu} x_\mu \overline{x_\nu} \geqq 0. \tag{24}$$

Conversely if $\rho(P)$ is a finite sum of the form

$$\sum_{p,m,\mu} x_{p,m\mu} \overline{\varphi_{p,m\mu}(P)} \tag{25}$$

the left side of (3) is

$$\sum_{p,m} \left(\sum_{\mu,\nu=1}^{h} a_{p,\mu\nu} x_{p,m\mu} \overline{x_{p,m\nu}} \right)$$

and this is $\geqq 0$ if (24) holds. But finite sums of the form (25) are dense in the family of all function $\rho(P)$ and thus (24) implies property 3). Finally by property (iii) of $\{\varphi(P)\}$, $f(P, Q)$ is real if and only if

$$\overline{a_{k,\nu\mu}} = a_{-k,\mu\nu}. \tag{26}$$

Thus a function $f(P, Q)$ of the real class $\mathfrak{P}$ has an expansion of the form

$$a_0 + \sum_{k=1}^{\infty} \left(\sum_{\mu,\nu=1}^{h(k)} a_{k,\mu\nu}(\varphi_{k,\mu\nu}(P, Q) + \overline{\varphi_{k,\mu\nu}(P, Q)}) \right) \tag{27}$$

where each matrix (23) is a semi-definite hermitian matrix. Before proceeding we require the following

LEMMA 2. *There exists an array of real numbers* $r_{n,k}$, $n = 1, 2, \cdots$, $k = 0, \pm 1, \pm 2, \cdots$, *with the following properties*: (i) $0 \leqq r_{n,k} \leqq 1$, (ii) $\lim_{n\to\infty} r_{n,k} = 1$, (iii) *for each n only a finite number of coefficients* $r_{n,k}$ *is* $\neq 0$, *and what is decisive,* (iv) *if* (21) *is the expansion of a continuous function* $f(P, Q)$, *then the sequence of functions*

$$f_n(P, Q) = \sum_{k=-\infty}^{\infty} r_{n,k} \left(\sum_{\mu,\nu=1}^{h} a_{k,\mu\nu}\varphi_{k,\mu\nu}(P, Q) \right) \tag{28}$$

is uniformly convergent towards $f(P, Q)$ *as* $n \to \infty$.

The proof of the lemma can be carried out along familiar lines and will be omitted.[5]

Now, if (23) is hermitian the number

$$\lambda_k = \sum_{\mu,\nu=1}^{h} a_{k,\mu\nu}\varphi_{k,\mu\nu}(\Phi, \Phi)$$

is $\geqq 0$ and

$$\left| \sum_{\mu,\nu=1}^{h} a_{k,\mu\nu}\varphi_{k,\mu\nu}(P, Q) \right| \leqq \lambda_k ;$$

therefore

$$0 \leqq \sum_{k=-\infty}^{\infty} r_{n,k}\lambda_k \leqq f_n(\Phi, \Phi) \leqq M$$

where M is independent of n. Letting $n \to \infty$ we obtain $\sum_k \lambda_k \leqq M$, and we hence conclude that the series (21) and (27) are absolutely and uniformly convergent. Therefore we have for

$$f(\Phi, \Phi) - f(P, Q) \equiv \tfrac{1}{2}f(P, P) + \tfrac{1}{2}f(Q, Q) - f(P, Q)$$

the series

$$\sum_{k=1}^{\infty} \left(\sum_{\mu,\nu=1}^{h(k)} a_{k,\mu\nu}\psi_{k,\mu\nu}(P, Q) \right) \tag{29}$$

where

$$\psi_{k,\mu\nu}(P, Q) = \sum_{m=1}^{l(k)} (\varphi_{k,m\mu}(P) - \varphi_{k,m\mu}(Q))(\overline{\varphi_{k,m\nu}(P)} - \overline{\varphi_{k,m\nu}(Q)}). \tag{30}$$

---

[5] See S. Bochner and J. v. Neumann, *Almost periodic functions in groups*, Trans. Amer. Math. Soc., 37 (1935), 21–50, esp. Part III; H. Weyl, loc. cit., p. 498–499.

Altogether, on the basis of Theorem 3 the following theorem can now be easily verified.

THEOREM 4. *A non-negative function $\rho(P, Q)$ is a group invariant Hilbert distance if and only if its square is an absolutely and uniformly convergent series of the form* (29) *in which such matrix* (23) *is non-negative hermitian and $\psi_{k,\mu\nu}$ has the value* (30).

We observe that our series (29) is a generalization of series (14).

THEOREM 5. *The Hilbert distance of Theorem* 4 *is certainly a proper distance if each matrix* (23) *is strictly positive definite.*

PROOF: In fact, if (23) is strictly positive then $\rho(P, Q)$ can only vanish if for each k, m, α

$$\varphi_{k,m\alpha}(P) - \varphi_{k,m\alpha}(Q) = 0.$$

But the latter fact implies that every continuous function on $\mathfrak{S}$ assumes equal values on the points P and Q, and this can only happen if P and Q are identical.

IV. SOME SPECIFIC CASES

Formulas (27) and (29) are greatly simplified if for all values of k we have $h(k) = 1$. In this case we can write

$$f(P, Q) = a_0 + \sum_{k=1}^{\infty} a_k(\varphi_k(P, Q) + \overline{\varphi_k(P, Q)})$$

where

$$\varphi_k(P, Q) + \overline{\varphi_k(P, Q)} = \sum_{m=1}^{l(k)} (\varphi_{k,m}(P)\overline{\varphi_{k,m}(Q)} + \overline{\varphi_{k,m}(P)}\varphi_{k,m}(Q)); \tag{31}$$

also the expression

$$\sum_{k=1}^{\infty} a_k(2\varphi_k(\Phi, \Phi) - \varphi_k(P, Q) - \overline{\varphi_k(P, Q)}) \tag{32}$$

for $\rho(P, Q)^2$ shows a strong resemblance to the left side of (16). The resemblance is most pronounced if $\mathfrak{S}$ is the $(m - 1)$-dimensional unit sphere in m-dimensional Euclidean space. In this case (31) is but for a numerical factor depending on k (and m) the expression

$$T_k^{(\lambda)}(\cos \vartheta)$$

where $T_k^{(\lambda)}$ is an ultraspherical polynomial; $\lambda = m - 2/2$, and ϑ is the geodesic distance between the points P, Q. The expression

$$\sum_{k=0}^{\infty} b_k T_k^{(\lambda)}(\cos \vartheta)$$

with $b_k \geqq 0$ and $\sum_k b_k T_k^{(\lambda)}(1) < \infty$, for the most general positive definite func-

tion depending on the geodesic distance alone, has been given previously by Schoenberg.[6]

More special than the assumption $h(k) = 1$ is the assumption $l(k) = 1$. It corresponds to the case of the group $\mathfrak{C}$ being an Abelian group, and in this case we simply have

$$f(P, Q) = \sum_{k=0}^{\infty} a_k(\varphi_k(P)\overline{\varphi_k(Q)} + \overline{\varphi_k(P)}\varphi_k(Q))$$

with $a_k \geqq 0$.

The symbol sP is a real multiplication of P by s if $\mathfrak{S}$ is a group space (in which case we write x, y, etc. instead of P, Q, etc.) and $\mathfrak{C}$ is its group of left transformations $x \rightarrow sx$. Our assumptions concerning $\mathfrak{S}$ are now that $\mathfrak{S}$ is a compact separable group, the invariant measure on it being now uniquely determined. In this case, $h(k) = l(k)$ and but for the factor $l(k)^{1/2}$, $\varphi_{k,\mu\nu}(x)$ is $u_{k,\mu\nu}(x)$. Since

$$\sum_{m=1}^{l} u_{k,m\mu}(x)\overline{u_{k,m\nu}(y)} = \sum_{m=1}^{l} u_{k,\nu m}(y^{-1})u_{k,m\mu}(x)$$

we see that, except for the factor $l(k)$,

$$\varphi_{k,\mu\nu}(x, y) = u_{k,\nu\mu}(y^{-1}x),$$

and hence we obtain the following result.

THEOREM 6. *If $\mathfrak{C}$ is a compact separable group, then $f(x, y)$ is a left invariant member of $\mathfrak{P}$ if and only if $f(x, y) = f(y^{-1}x)$ where*

$$f(t) = \sum_{k=-\infty}^{\infty} \sum_{\mu,\nu=1}^{l(k)} b_{k,\mu\nu} u_{k,\mu\nu}(t); \tag{33}$$

the matrix $| b_{k,\mu\nu} |$ being non-negative hermitian and the series (33) *in k being absolutely convergent.*

In order to obtain right invariant functions we have to put $t = yx^{-1}$. Our function is invariant on both sides if $f(s^{-1}y^{-1}xs) = f(y^{-1}x)$ that is if $f(t)$ is a class function. For class functions the expansion (33) depends only on the group characters

$$\chi_k(t) = \sum_{u=1}^{l} u_{k,\mu\mu}(t).^{7}$$

Thus we obtain

THEOREM 7. *If $\mathfrak{C}$ is a compact separable group, then $f(x, y)$ is a group invariant member of $\mathfrak{P}$, if and only if*

$$f(x, y) = \sum_{k=-\infty}^{\infty} b_k \chi_k(yx^{-1})$$

[6] I. J. Schoenberg, *On positive definite functions on spheres.* Bull. Amer. Math. Soc., 46 (1940), p. 888.

[7] See Bochner-von Neumann, loc. cit.

where $b_k \geqq 0$, *and* $\sum_k b_k \chi_k(1) < \infty$; *and* $\rho(x, y)$ *is a group invariant Hilbert distance if and only if*

$$\rho(x, y)^2 = \sum_{k=1}^{\infty} c_k \left(1 - \frac{\chi_k(yx^{-1}) + \chi_k(xy^{-1})}{2l(k)}\right)$$

with $c_k \geqq 0$; *the function* $\rho(x, y)$ *is certainly a proper distance if all* c_k *are* > 0.

V. Isometric Imbedding of Riemannian Space into Hilbert Space

In the present section we will draw a conclusion from Theorem 5. We assume that our space $\mathfrak{S}$ is a coordinate space of class C_r, $r \geqq 2$, (continuous partial derivatives of order $\leqq r$), or C_∞ (derivatives of every order) or C_ω (analytic coordinates), and we further assume explicitly that the functions $\{\varphi_{k,m_\mu}(P)\}$ belong to the same class. By choosing the coefficients of the matrix (23) sufficiently small we can obtain a series (29) whose sum will belong to the same class on $\mathfrak{S} \times \mathfrak{S}$. Now, by a general theorem,[8] such a distance can be generated by a Riemannian metric. Altogether we have the following

Theorem 8. *If* $\mathfrak{S}$ *is a compact differentiable manifold (of arbitrary dimension and) of class* C_r, $r \geqq 2$, *or* C_∞, *or* C_ω, *if* $\mathfrak{G}$ *is a fixed transitive group of homomorphisms on it, and if the corresponding generalized spherical harmonics* $\{\varphi_{k,m_\mu}(P)\}$ *belong to the same class, then there exists on* S *a group invariant positive definite Riemannian metric of class* C_{r-2}, *or* C_∞, *or* C_ω *for which the given space can be isometrically imbedded in real Hilbert space.*

Princeton University.

[8] S. Bochner, *Differentiable and Riemann metric*, Duke Math. Jour., 4 (1938), 51–54.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

UNITÄRINVARIANTE HYPERMAXIMALE NORMALE OPERATOREN

VON HIDEGORÔ NAKANO

(Received November 30, 1939)

Unitärinvariante hypermaximale Hermitesche Operatoren im Hilbertschen Raum sind schon in mehreren Formen gegeben.[1] Ich will in dieser Arbeit Unitärinvariante hypermaximale, normale Operatoren im Hilbertschen Raum $\mathfrak{H}$ in einer anderen Form geben.

Massoperatoren $E(Z)$ eines hypermaximalen, normalen Operators bilden einen abgeschlossenen, Abelschen Ring $\mathfrak{R}$ von Projektionsoperatoren, dessen Definition in der früheren Abhandlung[2] gegeben ist. Zuerst definieren wir die Dimension eines abgeschlossenen, Abelschen Rings. Für einen hypermaximalen, normalen Operator N kann man die komplexe Ebene G in eine Folge von Mengen Z_∞, Z_1, Z_2, $\cdots$ derart zerlegen, dass je zwei der Z_∞, Z_1, Z_2, $\cdots$ keinen gemeinsamen Punkt besitzen, und der aus allen Projektionsoperatoren $E(Z)$ $(Z \subset Z_n)$ gebildete, abgeschlossene, Abelsche Ring $\mathfrak{R}_n$ von der Dimension n ist. Diese Zerlegung ist bis auf eine Nullmenge über $E(Z)$ eindeutig bestimmt. Diese Zerlegung Z_∞, Z_1, Z_2, $\cdots$ nennen wir das Spektralsystem des hypermaximalen, normalen Operators N. Dafür, dass zwei hypermaximale, normale Operatoren N_1 und N_2 unitär äquivalent seien, ist es notwendig und hinreichend, dass die Klassen der Nullmengen und die Spektralsysteme von N_1 und N_2 miteinander übereinstimmen.

1. Dimension von Ring. Abgeschlossene Abelsche Ringe nennen wir kurz *Ringe. $\mathfrak{R}$ sei ein Ring, d.h. je zwei Projektionsoperatoren P und Q in $\mathfrak{R}$ sind vertauschbar, $P - PQ$ gehört zu $\mathfrak{R}$, und für jede Folge P_1, P_2, $\cdots$ in $\mathfrak{R}$ gehören $P_1 \dotplus P_2 \dotplus \cdots$ und $P_1P_2 \cdots$ beide zu $\mathfrak{R}$.*

DEFINITION 1. Ein Projektionsoperator K heisst mit einem Ring $\mathfrak{R}$ *vertauschbar*, wenn jedes P in $\mathfrak{R}$ mit K vertauschbar ist.

DEFINITION 2. Wenn ein Projektionsoperator K mit einem Ring $\mathfrak{R}$ vertauschbar ist, so bilden KP $(P \in \mathfrak{R})$ auch einen Ring, welchen wir den *Nebenring* von $\mathfrak{R}$ nennen.

Da $\mathfrak{H}$ separabel ist, kann man leicht beweisen, dass die Summe $\sum_{P \in \mathfrak{R}} \{P\}$ auch zu $\mathfrak{R}$ gehört.[3] Diese Summe nennen wir den *Maximaloperator* des Rings $\mathfrak{R}$.

[1] Vgl. M. H. Stone, *Linear Transformations in Hilbert Space*, New York (1932). F. Wecken, *Unitärinvarianten selbstadjungierter Operatoren*, Math. Annalen, 116 (1939).

[2] H. Nakano, *Über Abelsche Ringe von Projektionsoperatoren*, Proc. of the Phys.-Math. Soc. of Japan, 21 (1939).

[3] f_1, f_2, $\cdots$ sei überall dicht in $\mathfrak{H}$, und P_{n1}, P_{n2}, $\cdots$ sei eine Folge von Projektions-

Indem man den Maximaloperator von $\mathfrak{R}$ mit M bezeichnet, kann man leicht einsehen, dass zwei Nebenringe $K\mathfrak{R}$ und $L\mathfrak{R}$ dann und nur dann übereinstimmen, wenn $KM = LM$ ist, da der Ring $K\mathfrak{R}$ offenbar mit dem Ring $KM\mathfrak{R}$ übereinstimmt.

DEFINITION 3. Zwei Nebenringe $K\mathfrak{R}$ und $L\mathfrak{R}$ heissen *isomorph*, wenn für jeden Projektionsoperator P in $\mathfrak{R}$ $KP = 0$ mit $LP = 0$ gleiche Bedeutung hat.

DEFINITION 4. Zwei Nebenringe $K\mathfrak{R}$ und $L\mathfrak{R}$ heissen zueinander *orthogonal*, wenn für je zwei Projektionsoperatoren P und Q in $\mathfrak{R}$ stets $(KP)(LQ) = 0$ ist.

Man kann leicht einsehen, dass zwei Nebenringe $K\mathfrak{R}$ und $L\mathfrak{R}$ dann und nur dann zueinander orthogonal, sind wenn $(LM)(KM) = 0$ ist.

DEFINITION 5. Wenn n Nebenringe $K_1\mathfrak{R}, K_2\mathfrak{R}, \cdots, K_n\mathfrak{R}$ eines Rings $\mathfrak{R}$ alle mit $\mathfrak{R}$ isomorph, und je zwei von $K_1\mathfrak{R}, K_2\mathfrak{R}, \cdots, K_n\mathfrak{R}$ zueinander orthogonal sind, und noch $n + 1$ solche Nebenringe nicht existieren, so heisst n die *Minimaldimension* von $\mathfrak{R}$. Wenn beliebig viele solche Nebenringe existieren, so heisst $\mathfrak{R}$ *von der Dimension* ∞.

DEFINITION 6. $\mathfrak{R}$ sei ein Ring. Wenn für jedes P in $\mathfrak{R}$ die Minimaldimension des Nebenrings $P\mathfrak{R}$ stets n ist, so heisst $\mathfrak{R}$ *gleichmässig dimensional* mit der Dimension n.

Aus den Definitionen folgt sofort der folgende

SATZ 1. *Wenn die Minimaldimension eines Rings $\mathfrak{R}$ grösser als n ist, so ist die Minimaldimension des Nebenrings $P\mathfrak{R}$ für jedes P in $\mathfrak{R}$ auch grösser als n.*

SATZ 2. *Wenn die Minimaldimensionen von endlich oder abzählbar unendlich vielen Nebenringen $L_1\mathfrak{R}, L_2\mathfrak{R}, \cdots$ alle grösser als n sind, und $L_iL_j = 0$ $(i \neq j)$ gilt, so ist die Minimaldimension des Nebenrings $(L_1 + L_2 + \cdots)\mathfrak{R}$ auch grösser als n.*

BEWEIS: Nach Voraussetzung gibt es $n + 1$ zueinander orthogonale Nebenringe $K_i^{(1)}L_i\mathfrak{R}, K_i^{(2)}L_i\mathfrak{R}, \cdots, K_i^{(n+1)}L_i\mathfrak{R}$, welche alle mit $L_i\mathfrak{R}$ $(i = 1, 2, \cdots)$ isomorph sind. Dann sind $n + 1$ Nebenringe $(L_1K_1^{(1)} + L_2K_2^{(1)} + \cdots)\mathfrak{R}, \cdots, (L_1K_1^{(n+1)} + L_2K_2^{(n+1)} + \cdots)\mathfrak{R}$ offenbar zueinander orthogonal, und sie sind sogar alle mit $(L_1 + L_2 + \cdots)\mathfrak{R}$ isomorph, denn, aus $(L_1K_1^{(j)} + L_2K_2^{(j)} + \cdots)P = 0$ $(P \in \mathfrak{R})$ folgt $L_1K_1^{(j)}P = 0, L_2K_2^{(j)}P = 0, \cdots$, und weiter $L_1P = 0, L_2P = 0, \cdots$, folglich $(L_1 + L_2 + \cdots)P = 0$, und umgekehrt folgt aus $(L_1 + L_2 + \cdots)P = 0$ $(P \in \mathfrak{R})$ offenbar $(L_1K_1^{(j)} + L_2K_2^{(j)} + \cdots)P = 0$. Daher ist die Minimaldimension des Rings $(L_1 + L_2 + \cdots)\mathfrak{R}$ grösser als n.

SATZ 3. *Wenn die Minimaldimensionen von endlich oder abzählbar unendlich vielen Nebenringen $L_1\mathfrak{R}, L_2\mathfrak{R}, \cdots$ grösser als n sind, und je zwei von $L_1, L_2, \cdots$*

---

operatoren im Ring $\mathfrak{R}$, derart dass

$$\lim_{i\to\infty} \| P_{ni}f_n \| = \underset{R \in \mathfrak{R}}{\text{l.u.b.}} \| Rf_n \|$$

ist. Setzt man $P_n = \sum_{i,j \leq n} P_{ij}$, so ist $P_1 \leq P_2 \leq \cdots$, und alle P_n gehören zu $\mathfrak{R}$. Folglich gehört $\lim_{n\to\infty} P_n = M$ auch zu $\mathfrak{R}$, und es gilt $\| Mf_n \| = \underset{R \in \mathfrak{R}}{\text{l.u.b.}} \| Rf_n \|$ $(n = 1, 2, \cdots)$. Daher ist M der Maximaloperator in $\mathfrak{R}$.

miteinander vertauschbar sind, so ist die Minimaldimension des Nebenrings $(L_1 \dotplus L_2 \dotplus \cdots)\mathfrak{R}$ *grösser als* n.

Beweis: Setzt man

$$\bar{L}_n = L_n - L_n(L_1 \dotplus L_2 \dotplus \cdots \dotplus L_{n-1}),$$

so gilt

$$\bar{L}_i\bar{L}_j = 0 \qquad (i \neq j)$$

$$L_1 \dotplus L_2 \dotplus \cdots = \bar{L}_1 + \bar{L}_2 + \cdots,$$

und nach dem Satz 1 ist die Minimaldimension von $\bar{L}_n$ auch grösser als n. Daher folgt dieser Satz sofort aus dem Satz 3.

Satz 4. *Wenn ein Ring* $\mathfrak{R}$ *von der Minimaldimension* n *ist, so gibt es zwei Projektionsoperatoren* P *und* Q *in* $\mathfrak{R}$, *derart dass* $PQ = 0$, *der Nebenring* $P\mathfrak{R}$ *gleichmässig dimensional mit den Dimension* n, *die Minimaldimension des Nebenrings* $Q\mathfrak{R}$ *grösser als* n *ist, und* $P + Q$ *mit dem Maximaloperator* M *von* $\mathfrak{R}$ *übereinstimmt.*

Beweis: Wenn $\mathfrak{R}$ gleichmässig dimensional mit der Dimension n ist, so braucht man nur $P = M$ zu setzen. Im anderen Falle bilden alle Projektionsoperatoren R in $\mathfrak{R}$, für welche die Minimaldimension des Nebenrings $R\mathfrak{R}$ grösser als n ist, nach den Sätzen 1 und 3 einen Unterring $\mathfrak{R}_1$ von $\mathfrak{R}$. Bezeichnet man den Maximaloperator von $\mathfrak{R}_1$ mit Q, so gehört Q zu $\mathfrak{R}_1$, und die Minimaldimension des Nebenrings $Q\mathfrak{R}$ ist grösser als n, und für $P = M - Q$ ist der Nebenring $P\mathfrak{R}$ offenbar gleichmässig dimensional mit der Dimension n.

2. Unitärinvarianz. Es sei N ein hypermaximaler, normaler Operator mit dem Massoperator $E(Z)$. Wenn man mit $\mathfrak{R}$ den aus allen $E(Z)$ gebildeten Ring bezeichnet, so kann man nach dem Satz 4 die komplexe Ebene G in abzählbar unendlich viele, messbare Punktmengen Z_∞, Z_1, Z_2, $\cdots$ derart zerlegen, dass der Nebenring $E(Z_i)\mathfrak{R}$ gleichmässig dimensional mit der Dimension $i(= \infty, 1, \cdots)$ ist, und $G = Z_\infty + Z_1 + Z_2 + \cdots$, $Z_iZ_j = 0$ $(i \neq j)$ gilt. Diese Zerlegung Z_∞, Z_1, Z_2, $\cdots$ der komplexen Ebene G nennen wir das *Spektralsystem* des hypermaximalen, normalen Operators N. Man kann leicht einsehen, dass das Spektralsystem bis auf eine Nullmenge über $E(Z)$ eindeutig bestimmt ist.

Satz 5. *Dafür, dass zwei hypermaximale normale Operatoren* N_1 *und* N_2 *unitär äquivalent seien, d.h. dafür, dass für einen passenden unitären Operator* U $N_1 = U^*N_2U$ *bestehe, ist es notwendig und hinreichend, dass die Klassen der Nullmengen und die Spektralsysteme von* N_1 *und* N_2 *bis auf eine Nullmenge übereinstimmen.*

Mit $E_1(Z)$ und $E_2(Z)$ bezeichnen wir den Massoperator bzw. von N_1 und N_2. Wenn für einen unitären Operator U $N_1 = U^*N_2U$ ist, so erhält man nach der Eigenwertdarstellung

$$\int_G z\, d(E_1(Z)f, g) = \int_G z\, d(U^*E_2(Z)Uf, g)$$

für jedes f im Definitionsbereich von N_1 und jedes g in $\mathfrak{H}$, folglich nach der Eindeutigkeit des Massoperators

$$E_1(Z) = U^*E_2(Z)U.$$

Daher kann man leicht einsehen, dass die Klassen der Nullmengen über $E_1(Z)$ und $E_2(Z)$ und die Spektralsysteme von N_1 und N_2 bis auf eine Nullmenge übereinstimmen.

Für den Beweis der Hinreichendheit brauchen wir folgende Hilfssätze.

HILFSSATZ 1. *$\mathfrak{R}$ sei ein Ring mit dem Maximaloperator M. Es gibt mindestens ein Element f in $\mathfrak{H}$ derart, dass für jedes $R(\neq 0)$ in $\mathfrak{R}$ stets $Rf \neq 0$ ist. Ein solches Element f nennen wir ein eigentliches Element in bezug auf $\mathfrak{R}$.*

BEWEIS: Da $\mathfrak{H}$ separabel ist, gibt es eine Folge von Elementen $f_1, f_2, \cdots$ derart, dass $\mathfrak{H}$ von $f_1, f_2, \cdots$ aufgespannt wird, und $\|f_1\|^2 + \|f_2\|^2 + \cdots$ konvergiert. Alle Projektionsoperatoren R in $\mathfrak{R}$, für welche $Rf_n = 0$ ist, bilden zusammen einen Unterring $\mathfrak{R}_n$ von $\mathfrak{R}$. Wenn man den Maximaloperator von $\mathfrak{R}_n$ mit E_n bezeichnet, so kann man leicht einsehen, dass $\mathfrak{R}_n$ mit dem Nebenring $E_n\mathfrak{R}$ übereinstimmt. Daher ist $(M - E_n)Rf_n \neq 0$ für jedes R in $\mathfrak{R}$ wenn $(M - E_n)R \neq 0$ ist. Setzt man $E'_n = E_1E_2 \cdots E_n$, so konvergiert

$$(M - E'_1)f_1 + (E'_1 - E'_2)f_2 + (E'_2 - E'_3)f_3 + \cdots$$

nach einem Element f_0, da $\|(E'_{n-1} - E'_n)f_3\|^2 \leqq \|f_3\|^2$ ist, und $\|f_1\|^2 + \|f_2\|^2 + \cdots$ konvergiert. Dieses Element f_0 ist ein eigentliches Element in bezug auf $\mathfrak{R}$, denn, wenn für einen R in $\mathfrak{R}$ $Rf_0 = 0$ gilt, dann ist zugleich

$$(M - E'_1)Rf_0 = 0, \qquad (E'_1 - E'_2)Rf_0 = 0, \qquad (E'_2 - E'_3)Rf_0 = 0, \cdots,$$

folglich muss

$$(M - E'_1)R = 0, \quad (M - E_2)(E'_1 - E'_2)R = 0, \quad (M - E_3)(E'_2 - E'_3)R = 0, \cdots$$

sein. Da

$$\begin{aligned}(M - E_n)(E'_{n-1} - E'_n) &= (M - E_n)(E_1 \cdots E_{n-1} - E_1 \cdots E_{n-1}E_n)\\ &= E_1 \cdots E_{n-1} - E_1 \cdots E_{n-1}E_n = E'_{n-1} - E'_n\end{aligned}$$

ist, erhält man

$$(M - E'_n)R = 0, \qquad \text{d.h.} \qquad MR = E'_nR$$

folglich

$$MR = R \lim_{n\to\infty} E'_n.$$

Andererseits gilt $(\lim_{n\to\infty} E'_n)f_i = 0$ $(i = 1, 2, \cdots)$, und $f_1, f_2, f_3, \cdots$ spannen $\mathfrak{H}$ auf. Daher muss $\lim_{n\to\infty} E'_n = 0$ sein, folglich $R = MR = R(\lim_{n\to\infty} E'_n) = 0$.

DEFINITION 7. Zwei in bezug auf $\mathfrak{R}$ eigentliche Elemente f und g heissen *eigentlich orthgonal*, wenn für jedes R in $\mathfrak{R}(Rf, g) = (f, Rg) = 0$ gilt.

Definition 8. $\mathfrak{R}$ sei ein Ring mit dem Maximaloperator M. Wenn für n eigentliche Elemente $f_1, f_2, \cdots, f_n$ ($n = \infty, 1, 2, \cdots$), von denen je zwei zueinander eigentlich orthogonal sind, die abgeschlossene lineare Mannigfaltigkeit des Maximaloperators aus allen $\{Rf_1, Rf_2, \cdots, Rf_n\}$ ($R \in \mathfrak{R}$) aufgespannt wird, so heisst $f_1, f_2, \cdots, f_n$ ein *Basensystem* des Rings $\mathfrak{R}$.

Hilfssatz 2. *Wenn ein Ring $\mathfrak{R}$ mit dem Maximaloperator M gleichmässig dimensional mit endlicher Dimension n ist, so besitzt $\mathfrak{R}$ ein Basensystem, das aus n eigentlichen Elementen besteht.*

Beweis: Da die Dimension von $\mathfrak{R}$ n ist, so gibt es zueinander orthogonale, mit $\mathfrak{R}$ isomorphe Nebenringe $K_1\mathfrak{R}$, K_2R, $\cdots$, $K_n\mathfrak{R}$ mit den Maximaloperatoren $K_1, K_2, \cdots, K_n$. Wenn man mit f_i ein eigentliches Element in bezug auf $K_i\mathfrak{R}$ bezeichnet, so sind je zwei der K_1f_1, K_2f_2, $\cdots$, K_nf_n offenbar eigentlich orthogonal. und K_if_i ist ein eigentliches Element in bezug auf $K_i\mathfrak{R}$. Bezeichnet man mit $\bar{K}_i$ den Projektionsoperator der von allen $\{RK_if_i\}$ ($R \in \mathfrak{R}$) aufgespannten, abgeschlossenen, linearen Mannigfaltigkeit, so ist offenbar $M \geqq \bar{K}_1 + \bar{K}_2 + \cdots + \bar{K}_n$ .Wenn $M - (\bar{K}_1 + \bar{K}_2 + \cdots + \bar{K}_n) \neq 0$ ist, so bilden alle Projektionsoperatoren R in $\mathfrak{R}$, für die $[M - (\bar{K}_1 + \bar{K}_2 + \cdots + \bar{K}_n)]R = 0$ ist, einen Ring mit dem Maximaloperator R_0, und es ist $M - R_0 \neq 0$. Dann ist die Minimaldimension von $(M - R_0)\mathfrak{R}$ grösser als n, was unmöglich ist, da $\mathfrak{R}$ gleichmässig dimensional mit der Dimension n ist. Daher ist $M = \bar{K}_1 + \bar{K}_2 + \cdots + \bar{K}_n$ d.h. $f_1, f_2, \cdots, f_n$ ist ein Basensystem von $\mathfrak{R}$.

Hilfssatz 3. *Wenn ein Ring $\mathfrak{R}$ mit dem Maximaloperator M ein aus n Elementen bestehendes Basensystem $f_1, f_2, \cdots, f_n$ besitzt, so ist $\mathfrak{R}$ von der Minimaldimension n.*

Beweis: Wenn man mit $\mathfrak{M}$ die dem Projektionsoperator $\mathfrak{M}$ entsprechende, abgeschlossene, lineare Mannigfaltigkeit bezeichnet, so gibt es einen hypermaximalen, normalen Operator N in $\mathfrak{M}$, dessen Eigenring mit $\mathfrak{R}$ übereinstimmt.[4] Für jedes f_0 in $\mathfrak{M}$ lässt sich jedes Element in der aus allen Rf_0($R \in \mathfrak{R}$) aufgespannten, abgeschlossenen, linearen Mannigfaltikeit in der Form $\Phi(N)f_0$ darstellen, wo $\Phi(N)$ eine Funktion von N, und $\Phi(N)f_0$ sinnvoll ist.[5] Da Mf_1, Mf_2, $\cdots$, Mf_n auch ein Basensystem von $\mathfrak{R}$ ist, so kann man annehmen, dass $f_1, f_2, \cdots, f_n$ alle zu $\mathfrak{M}$ gehören, und wir betrachten daher nur den Raum $\mathfrak{M}$. Wenn die

[4] Satz 8 in der Abhandlung: H. Nakano, *Über Abelsche Ringe von Projektionsoperatoren*, Proc. of the Phys.-Math. Soc. of Japan 21 (1939).

[5] Vgl. M. H. Stone, *Linear Transformations in Hilbert Space*, New York (1932). Man kann es auch wie folgt einsehen: Alle über $E(Z)$ messbare Funktionen $\Phi(z)$, für die das Integral $\int_G |\Phi(z)|^2 d\, \|E(Z)f_0\|^2$ konvergiert, bilden nach dem Riesz-Fisherschen Satz einen Hilbertschen Raum mit dem inneren Produkt $\int_G \Phi_1(z)\overline{\Phi_1(z)}\, d\, \|E(Z)f_0\|^2$. Dieser Hilbertsche Raum ist mit der aus allen $\Phi(N)f_0$ bestehenden, linearen Mannigfaltigkeit isometrisch, da $(\Phi_1(N)f_0, \Phi_2(N)f_0) = \int_G \Phi_1(z)\overline{\Phi_2(z)}\, d\, \|E(Z)f_0\|^2$ ist. Daher bilden $\Phi(N)f_0$ eine abgeschlossene lineare Mannigfaltigkeit, die offenbar von allen $E(Z)f_0$ aufgespannt wird.

Minimaldimension von $\mathfrak{R}$ grösser als n ist, so kann man wie im obigen Beweis leicht einsehen, dass zueinander eigentlich orthogonale Elemente $\varphi_1, \varphi_2, \cdots, \varphi_m$ $(m > n)$ existieren. Da $\mathfrak{M}$ von allen $\{Rf_1, Rf_2, \cdots, Rf_n\}(R \in \mathfrak{R})$ aufgespannt wird, so kann man schreiben:

$$(*) \qquad \begin{array}{c} \varphi_1 = \Phi_{11}f_1 + \cdots + \Phi_{1n}f_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \varphi_m = \Phi_{m1}f_1 + \cdots + \Phi_{mn}f_n \end{array} \qquad (m > n)$$

wo Φ_{ij} passende Funktionen von N sind, daher sind Φ_{ij} alle hypermaximal, ihr Eigenring ist in $\mathfrak{R}$ enthalten, und je zwei der $\Phi_{ij}\begin{pmatrix} i = 1, 2, \cdots, m \\ j = 1, 2, \cdots, n \end{pmatrix}$ sind miteinander vertauschbar. Da Φ_{ij} hypermaximal ist, so gibt es für jede positive Zahl p einen Projektionsoperator $R_{ij}^{(p)}$ in $\mathfrak{R}$ derart, dass $\| \Phi_{ij}R_{ij}^{(p)}f \| \leqq p \, \| R_{ij}^{(p)}f \|$ für jedes f in $\mathfrak{M}$, und $\lim_{p\to\infty} R_{ij}^{(p)} = M$ ist. Setzt man $R^{(p)} = \prod_{ij} R_{ij}^{(p)}$, so gilt $\lim_{p\to\infty} R^{(p)} = M$. Daher ist $R^{(p)} \neq 0$ für eine genügend grosse, positive Zahl p, und $\Phi_{ij}R^{(p)}$ sind alle beschränkt. Aus (*) erhält man

$$\begin{array}{c} R^{(p)}\varphi_1 = \Phi_{11}R^{(p)}f_1 + \cdots + \Phi_{1n}R^{(p)}f_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ R^{(p)}\varphi_m = \Phi_{m1}R^{(p)}f_1 + \cdots + \Phi_{mn}R^{(p)}f_n \end{array} \qquad (m > n)$$

wobei $R^{(p)}\varphi_i \neq 0$ $(i = 1, 2, \cdots, m)$ ist, da alle $\varphi_1, \varphi_2, \cdots, \varphi_m$ eigentliche Elemente in bezug auf $\mathfrak{R}$ sind. Daher verschwinden nicht alle Operatoren $\Phi_{ij}R^{(p)}$ identisch in $\mathfrak{M}$. Nun kann man wie in der Algebra beschränkte Operatoren Ψ_1, $\Psi_2, \cdots, \Psi_m$ in der Form von Determinanten mit den Elementen $\Phi_{ij}R^{(p)}$ derart bestimmen, dass

$$\Psi_1 R^{(p)}\varphi_1 + \Psi_2 R^{(p)}\varphi_2 + \cdots + \Psi_m R^{(p)}\varphi_m = 0$$

ist, und nicht alle $\Psi_1 R^{(p)}, \Psi_2 R^{(p)}, \cdots, \Psi_m R^{(p)}$ identisch in $\mathfrak{M}$ verschwinden. Da je zwei der $\varphi_1, \varphi_2, \cdots, \varphi_m$ zueinander eigentlich orthogonal sind, so kann man durch Eigenwertdarstellung leicht einsehen, dass je zwei der $\Psi_1 R^{(p)}\varphi_1$, $\Psi_2 R^{(p)}\varphi_2, \cdots, \Psi_m R^{(p)}\varphi_m$ zueinander orthogonal sind. Daher muss

$$\Psi_1 R^{(p)}\varphi_1 = 0, \qquad \Psi_2 R^{(p)}\varphi_2 = 0, \cdots, \Psi_m R^{(p)}\varphi_m = 0$$

sein. Da nicht alle Operatoren $\Psi_1 R^{(p)}, \Psi_2 R^{(p)}, \cdots, \Psi_m R^{(p)}$ identisch verschwinden, kann man annehmen, dass $\Psi_1 R^{(p)}$ nicht identisch verschwindet. Dann gibt es einen Projektionsoperator R_1 $(\neq 0)$ in $\mathfrak{R}$, sodass für jedes f in $\mathfrak{M}$ aus $R_1 f \neq 0$, $\Psi_1 R^{(p)}Rf \neq 0$ folgt, da $\Psi_1 R^{(p)}$ auch eine Funktion von N ist. Für dieses R_1 gilt $\Psi_1 R^{(p)}R_1\varphi_1 = 0$, folglich $R_1\varphi_1 = 0$, was unmöglich ist, weil φ_1 ein eigentliches Element in bezug auf $\mathfrak{R}$ ist.

Hilfssatz 4. *Ein Ring $\mathfrak{R}$ mit der Dimension ∞ hat ein Basensystem f_1, $f_2, \cdots$.*

Beweis: $f_1, f_2, \cdots$ seien in $\mathfrak{H}$ überall dicht. Alle Projektionsoperatoren R in $\mathfrak{R}$, für welche $Rf_1 = 0$ ist, bilden einen Ring mit dem Maximaloperator R_1.

Es gibt nach dem Hilfssatz 2 ein eigentliches Element φ_1 in bezug auf dem Nebenring $R_1 \mathfrak{R}$. $g_1 = f_1 + R_1\varphi_1$ ist dann ein eigentliches Element in bezug auf $\mathfrak{R}$, denn, da $(Rf_1, RR_1\varphi_1) = (RR_1f_1, \varphi_1) = 0$ ist, folgt aus $R(f_1 + R_1\varphi_1) = 0$ für ein R in $\mathfrak{R} Rf_1 = 0$, $RR_1\varphi_1 = 0$, und weiter $RR_1 = R$, $RR_1 = 0$, d. h. $R = 0$.

Bezeichnet man mit K_1 den Projektionsoperator der von allen $\{R(f_1 + R_1\varphi_1)\}$ $(R \in \mathfrak{R})$ aufgespannten, abgeschlossenen, linearen Mannigfaltigkeit, so ist der Nebenring $(1 - K_1)\mathfrak{R}$ mit $\mathfrak{R}$ isomorph, denn, wäre $(1 - K_1)R = 0$ für ein $R(\neq 0)$ in $\mathfrak{R}$, so müsste die Minimaldimension des Nebenrings $K_1R\mathfrak{R}$ gleich 1 sein, da $f_1 + R_1\varphi_1$ ein Basensystem von $K_1R\mathfrak{R}$ ist, was nach dem Satz 1 unmöglich ist.

Alle Projektionsoperatoren R in $\mathfrak{R}$, für welche $R(1 - K_1)f_2 = 0$ ist, bilden einen Ring mit dem Maximaloperator R_2. Es gibt nach dem Hilfssatz 2 ein eigentliches Element φ_2 in bezug auf dem Nebenring $R_2\mathfrak{R}$. $g_2 = (1 - K_1)(f_2 + R_2\varphi_2)$ ist dann ein eigentliches Element in bezug auf $\mathfrak{R}$, das zu $f_1 + R_1\varphi_1$ eigentlich orthogonal ist.

Bezeichnet man mit K_2 den Projektionsoperator der aus allen $\{R(1 - K_1)g_2\}$ $(R \in \mathfrak{R})$ aufgespannten, abgeschlossenen, linearen Mannigfaltigkeit, so kann man auch wie oben schliessen, dass $[1 - (K_1 + K_2)]\mathfrak{R}$ mit $\mathfrak{R}$ isomorph ist. Ähnlich erhält man eine Folge von eigentlichen Elementen $g_1, g_2, \cdots$, von denen je zwei zueinander eigentlich orthogonal sind. Bezeichnet man mit M den Maximaloperator von $\mathfrak{R}$, so gehören alle $Mf_1, Mf_2, \cdots, Mf_n$ zu der von allen $\{Rg_1, Rg_2, \cdots, Rg_n\}(R \in \mathfrak{R})$ aufgespannten, abgeschlossenen, linearen Mannigfaltigkeit. Daher ist $g_1, g_2, \cdots$ ein Basensystem des Rings $\mathfrak{R}$.

Beweis des Satzes 5. N_1 und N_2 seien zwei hypermaximale, normale Operatoren mit den Massoperatoren. $E_1(Z)$ und $E_2(Z)$. Jede Nullmenge über einem der $E_i(Z)$ sei auch Nullmenge über dem anderen, und Z_∞, $Z_1, Z_2, \cdots$ sei ein gemeinsames Spektralsystem von N_1 und N_2. Wenn man die Eigenringe von N_1 und N_2 bzw. mit $\mathfrak{R}_1$ und $\mathfrak{R}_2$ bezeichnet, so besteht der Nebenring $E_1(Z_n)\mathfrak{R}_1$ aus allen $\{E_1(Z)\}(Z \subset Z_n)$, und der Nebenring $E_2(Z_n)\mathfrak{R}_n$ aus allen $\{E_2(Z)\}(Z \subset Z_n)(n = \infty, 1, 2, \cdots)$. $f_{n,1}, f_{n,2}, \cdots, f_{n,n}$ und $g_{n,1}$, $g_{n,2}, \cdots, g_{n,n}$ seien Basensysteme bzw. von $E_1(Z_n)\mathfrak{R}_1$ und $E_2(Z_n)\mathfrak{R}_2$ $(n = \infty, 1, 2, \cdots)$. Für beliebige beschränkte Funktionen

$$\Phi_{ij}(N_i) \begin{pmatrix} i = \infty, 1, 2, \cdots & i \geqq j \\ j = \quad 1, 2, 3, \cdots & \end{pmatrix}$$

sind dann je zwei der $\Phi_{ij}(N_1)E_i(Z_i)f_{i,j}$ zueinander orthogonal. Folglich bilden alle endlichen Summen aus ihnen eine lineare Mannigfaltigkeit $\mathfrak{N}_1$. $\mathfrak{N}_1$ ist überall dicht in $\mathfrak{H}$, da Massoperator $E_1(Z)$ auch eine beschränkte Funktion von N_1 ist. Die zwei Mengenfunktionen $\| E_1(Z)E_1(Z_n)f_{n,j} \|^2$ und $\| E_2(Z)E_2(Z_n)g_{n,j} \|^2$ verschwinden für eine Menge Z beide zugleich, wenn eine von ihnen verschwindet. Daher gibt es eine messbare Funktion $\Psi_{n,j}(z)$, so dass

$$\| E_1(Z)E_1(Z_n)f_{n,j} \|^2 = \int_G | \Psi_{nj}(z) |^2 d \| E_2(Z)E_2(Z_n)g_{n,j} \|^2$$
$$\begin{pmatrix} n = \infty, 1, 2, \cdots & n \geqq j \\ j = \quad 1, 2, 3, \cdots & \end{pmatrix}$$

ist.[6] Dann ist $f'_{n,j} = \Psi_{n,j}(N_2)E_2(Z_n)g_{n,j}$ $(j = 1, 2, \cdots, n)$ auch ein Basensystem von $E_2(Z_n)\mathfrak{R}_2$, und es gilt

$$\| E_1(Z)E_1(Z_n)f_{n,j} \| = \| E_2(Z)E_2(Z_n)f'_{n,j} \|.$$

Wenn man jeder beliebigen, endlichen Summe $f_1 = \sum_{n,j} \Phi_{nj}(N_1)E_1(Z_n)f_{n,j}$ die entsprechende Summe $f_2 = \sum_{n,j} \Phi_{nj}(N_2)E_2(Z_n)f'_{n,j}$ zuordnet, so ist diese Zuordnung längentreu, da

$$\begin{aligned} \|f_1\|^2 &= \| \sum_{n,j} \Phi_{nj}(N_1)E_1(Z_n)f_{n,j} \|^2 = \sum_{n,j} \| \Phi_{nj}(N_1)E_1(Z_n)f_{n,j} \|^2 \\ &= \sum_{n,j} \int_{Z_n} | \Phi_{nj}(z) |^2 d \| E_1(Z)E_1(Z_n)f_{n,j} \|^2 \\ &= \sum_{n,j} \int_{Z_n} | \Phi_{nj}(z) |^2 d \| E_2(Z)E_2(Z_n)f'_{n,j} \|^2 \\ &= \| \sum_{n,j} \Phi_{nj}(N_2)E_2(Z_n)f'_{n,j} \|^2 = \|f_2\|^2. \end{aligned}$$

Die $\sum_{n,j} \Phi_{nj}(N_2)E_2(Z_n)f'_{n,j}$ sind überall dicht in $\mathfrak{H}$. Daher kann man diese Zuordnung zu einer unitären Zuordnung $f_2 = Uf_1$ eindentig erweitern. Für jede charakteristische Funktion $\Lambda(z)$ einer messbaren Menge Z ist $E_1(Z) = \Lambda(N_1)$, $E_2(Z) = \Lambda(N_2)$. Daher erhält man

$$\begin{aligned} UE_1(Z)\{\sum_{n,j} \Phi_{nj}(N_1)E_1(Z_n)f_{n,j}\} &= U\}\sum_{n,j} \Phi_{nj}(N_1)\Lambda(N_1)E_1(Z_n)f_{n,j}\} \\ &= \sum_{n,j} \Phi_{nj}(N_2)\Lambda(N_2)E_2(Z_n)f'_{n,j} = E_2(Z)\{\sum_{n,j} \Phi_{nj}(N_2)E_2(Z_n)f'_{n,j}\} \\ &= E_2(Z)U\{\sum_{n,j} \Phi_{nj}(N_1)E_1(Z_n)f_{n,j}\}. \end{aligned}$$

Folglich gilt für jede messbare Menge Z

$$E_1(Z) = U^*E_2(Z)U,$$

und hieraus kann man durch Eigenwertdarstellung leicht schliessen, dass $N_1 = U^*N_2U$ ist.

IMPERIAL UNIVERSITY, TOKYO, JAPAN.

[6] Radon-Nykodymscher Satz. Vgl. S. Saks, *Theory of the Integral*, Warszawa (1937), S. 36.

Annals of Mathematics
Vol. 42, No. 3, July, 1941

ÜBER DEN BEWEIS DES STONESCHEN SATZES

Von Hidegorô Nakano

(Received February 13, 1940)

Der wohlbekannte Stonesche[1]

Satz. *$U_t(-\infty < t < +\infty)$ sei eine Schar unitärer Operatoren im Hilbertschen Raum $\mathfrak{H}$. Wenn stets*

$$U_t U_s = U_{t+s}$$

gilt, und für jedes f und g in $\mathfrak{H}$ $(U_t f, g)$ als eine Funktion von t messbar ist, so gibt es eine Zerlegung der Einheit $E(\lambda)$, derart dass

$$U_t = \int_{-\infty}^{+\infty} e^{it\lambda}\, dE(\lambda).$$

wurde von J. von Neumann,[2] M. H. Stone,[3] S. Bochner,[4] F. Riesz[5] und B. von Sz. Nagy[6] bewiesen. Unter diesen Beweisen beruht nur der von Neumannsche auf keinem Satz über Fouriersche Integrale oder Fouriersche Reihen. In dieser Abhandlung soll der von Neumannsche Beweis wesentlich vereinfacht werden—insbesondere wird ein besonderer Beweis der Stetigkeit von U_t in t und die damit zusammenhängende Bezugnahme auf die reelle Funktionentheorie vermieden.

Beweis. Wie von Neumann Zeigt, kann man durch

$$A = \int_0^{+\infty} e^{-t} U_t\, dt \tag{1}$$

einen beschränkten Operator erhalten. Es gilt dann weiter

$$A^* = \int_{-\infty}^{0} e^{t} U_t\, dt \tag{2}$$

$$U_s A = A U_s = e^s \int_s^{+\infty} e^{-t} U_t\, dt \tag{3}$$

$$AA^* = A^*A, \qquad A + A^* = 2AA^*. \tag{4}$$

Ferner gilt:

$$\text{aus} \quad Af = 0 \quad \text{folgt} \quad f = 0. \tag{5}$$

[1] M. H. Stone, Proc. Nat. Ac. 16, February 1930. S. 173–174.
[2] J. von Neumann, Annals of Math. 33, 1932. S. 567–573.
[3] M. H. Stone, Annals of Math. 33, 1932. S. 643–648.
[4] S. Bochner, Sitz. Preusz. Akad. 1933.
[5] F. Riesz, Acta Sci. Math. Szeged, 6, 1934.
[6] B. von Sz. Nagy, Math. Annalen 112, 1936. S. 291–294.

Denn wegen (3) folgt für jedes g

$$\int_s^{+\infty} e^{-t}(U_t f, g)\, dt = 0.$$

Da s beliebig ist, gilt $(U_t f, g) = 0$ bis auf eine Nullmenge in t. Dann kann man nach Separabilität von $\mathfrak{H}$ leicht beweisen, dass bis auf eine Nullmenge in t für alle g gleichzeitig $(U_t f, g) = 0$ d.h. $U_t f = 0$, folglich $\| f \| = \| U_t f \| = 0$.

Sei nun

$$W = -2A + 1. \tag{6}$$

Dann folgt aus (4) $W^*W = WW^* = 1$, d.h. W ist unitär. Und wegen (5) folgt aus $Wf = f, f = 0$. Daher ist W nach J. von Neumann,[7] S. 91, 92 die Cayleysche Transformierte eines hypermaximalen Hermiteschen Operators

$$H = \int_{-\infty}^{+\infty} \lambda\, dE(\lambda),$$

für eine geeignete Zerlegung der Einheit $E(\lambda)$. Die ebendort durchgeführten Rechnungen ergeben weiterhin

$$W = \int_{-\infty}^{+\infty} \frac{\lambda - i}{\lambda + i}\, dE(\lambda).$$

Im Hinblick auf (6) folgt hieraus

$$A = \int_{-\infty}^{+\infty} \frac{1}{1 - i\lambda}\, dE(\lambda). \tag{7}$$

Setzt man

$$V_t = \int_{-\infty}^{+\infty} e^{it\lambda}\, dE(\lambda)$$

für diese Zerlegung der Einheit $E(\lambda)$, so ist V_t unitär und stetig in t, und es gilt

$$V_{t+s} = V_t V_s \tag{8}$$

$$A = \int_0^{+\infty} e^{-t}\, V_t\, dt \tag{9}$$

$$A^* = \int_{-\infty}^{0} e^{t}\, V_t\, dt \tag{10}$$

$$V_s A = A V_s = e^s \int_s^{+\infty} e^{-t}\, V_t\, dt. \tag{11}$$

Setzt man für eine positive Zahl $n(>1)$ $E_n = E(n) - E(-n)$, und bezeichnet den Wertvorrat von E_n mit $\mathfrak{M}_n$, so gilt für jedes f in $\mathfrak{M}_n$ nach (7)

$$A = \int_{-n}^{n} \frac{1}{1 - i\lambda}\, dE(\lambda).$$

---

[7] J. von Neumann, Math. Annalen 102, 1929. S. 49–131.

Infolgedessen definiert in $\mathfrak{M}_n$

$$A^{-1} = \int_{-n}^{n} (1 - i\lambda)\, dE(\lambda)$$

eine Reciproke A^{-1} von A, die ausserdem beschränkt ist: Für alle f in $\mathfrak{M}_n$ ist

$$\| A^{-1}f \| \leqq 2n \| f \|, \qquad \| Af \| \geqq \frac{1}{2n} \| f \|.$$

(1), (3), (9), (11) verbindend kann man auch schreiben, dass

$$AU_s = e^s A - e^s \int_0^s e^{-t} U_t\, dt$$

$$AV_s = e^s A - e^s \int_0^s e^{-t} V_t\, dt,$$

daher

$$A(U_s - V_s) = -e^s \int_0^s e^{-t}(U_t - V_t)\, dt.$$

Da die abgeschlossene lineare Mannigfaltigkeit $\mathfrak{M}_n$ beide unitäre Operatoren U_s und V_s reduziert, so gilt für jede positive Zahl s

$$\frac{1}{2n} \| (U_s - V_s)f \| \leqq \| A(U_s - V_s)f \| \leqq e^s \int_0^s e^{-t} \| (U_t - V_t)f \|\, dt.$$ [8]

Setzt man $\varphi(s) = e^{-s} \| (U_s - V_s)f \|$, so ist

$$0 \leqq \frac{1}{2n} \varphi(s) \leqq \int_0^s \varphi(t)\, dt, \tag{12}$$

folglich

$$\frac{1}{2n} \int_0^s e^{-2ns} \varphi(s)\, ds \leqq \int_0^s \left\{ e^{-2ns} \left[\int_0^s \varphi(t)\, dt \right] \right\} ds$$
$$= -\frac{1}{2n} e^{-2ns} \int_0^s \varphi(t)\, dt + \frac{1}{2n} \int_0^s e^{-2ns} \varphi(s)\, ds,$$

und somit

$$\frac{1}{2n} e^{-2ns} \int_0^s \varphi(t)\, dt \leqq 0.$$

Daher muss nach (12) $\varphi(s) = 0$ sein, d.h. $U_s f = V_s f$ für jedes f in $\mathfrak{M}_n$. Für beliebiges f in $\mathfrak{H}$ gilt auch $U_s E_n f = V_s E_n f$. Da $\lim_{n \to \infty} E_n = 1$ ist, erhält man $U_s f = V_s f$, was zu beweisen war.

Hier danke ich Herrn Prof. J. von Neumann für freundliche Verbesserungen und Bemerkungen.

---

[8] Wenn $(Af, g) = \int (W_t f, g)\, dt$ ist, so gilt

$$| (Af, g) | \leqq \left| \int | (W_t f, g) |\, dt \right| \leqq \left| \int \| W_t f \| \cdot \| g \|\, dt \right| \leqq \| g \| \left| \int \| W_t f \|\, dt \right|.$$

Setzt man $g = Af$, so erhält man

$$\| Af \| \leqq \left| \int \| W_t f \|\, dt \right|.$$

Annals of Mathematics
Vol. 42, No. 3, July, 1941

ON THE TOPOLOGICAL STRUCTURE OF SOLVABLE GROUPS

By Claude Chevalley

(Received November 1, 1940)

E. Cartan has shown that a simply connected solvable Lie group is homeomorphic to some cartesian space.[1] We want to investigate the structure of a solvable Lie group which is not necessarily simply connected.

From a well known theorem,[2] it follows that such a group may be considered as the factor group of a solvable simply connected group G by a discrete sub-group D of the center of G. Therefore what we have to do is to look for the possible "situation" of such a sub-group D in the group G.

I. Notations. Generalities

G being a Lie group, to every element L of its Lie algebra $\mathfrak{L}$ there corresponds a one parametric sub-group $\sigma = \sigma(t)$ of G: we have $\sigma(t_1 + t_2) = \sigma(t_1)\sigma(t_2)$, and L is the tangent vector at the unit element to the differentiable curve described by $\sigma(t)$. Moreover, replacing L by λL, where λ is a real number, does not change the geometric curve $\sigma(t)$, but replaces the parameter t by $\frac{1}{\lambda} t$.

Extending to general Lie groups a notion which is well known for matrices, we shall denote by exp L the point of parameter 1 on the curve $\sigma(t)$. Therefore we have $(\exp tL)(\exp t'L) = \exp (t + t')L$.

If we transform the elements of the group $\exp Lt$ by a fixed element σ, we obtain a new one-parametric group $\exp L't$; the mapping $L \to L'$ is a linear mapping A_σ of the Lie algebra $\mathfrak{L}$ into itself; the mapping $\sigma \to \mathsf{A}_\sigma$ is a linear representation of the group G, which is called the *adjoint representation.* Any central element of G is obviously represented by the identity operation.

Suppose that σ describes itself a new one parametric group $\exp Mu$. Then L' becomes a function of u, and we have $L' = L + u[M, L] + \cdots$, the terms not written being of order > 1.

The mapping $L \to [M, L]$, for M fixed, is a linear operation A_M in $\mathfrak{L}$; the mapping $M \to \mathsf{A}_M$ is called the adjoint representation[3] of the Lie algebra; we have

$$\mathsf{A}_{[M,M']} = \mathsf{A}_{M'}\mathsf{A}_M - \mathsf{A}_M\mathsf{A}_{M'}.$$

[1] It follows at once from the lemma proved by *E. Cartan* on page 20 of the pamphlet "La topologie des groupes de Lie" (Exposes de geometrie, Act. scient. et indus. no. 358, Hermann et Cie, Paris, 1936). See also Theorem 83, p. 277 in *Pontrjagin,* "Topological groups" (Princeton University Press, 1939), Theorem 61, p. 225.

[2] See *Pontrjagin,* l. c. 1), Theorem 83, p. 277.

[3] Our notations do not coincide entirely with the usual ones. It is the mapping $M \to -A_M$ which is generally called the adjoint representation.

Let H be a closed invariant sub-group of G; it is also a Lie group, and its Lie algebra $\mathfrak{H}$ is an invariant sub-algebra—or ideal—in $\mathfrak{L}$. Moreover G/H is also a Lie group, whose Lie algebra is the quotient algebra $\mathfrak{L}/\mathfrak{H}$. If L, L' are two elements of $\mathfrak{L}$, and if L is a congruent to L' modulo $\mathfrak{H}$, i.e. if $L' - L \in \mathfrak{H}$ then the elements $\exp L$, $\exp L'$ belong to the same coset of G modulo H.

II. Solvable Groups

A Lie algebra $\mathfrak{L}$ is *solvable*[4] when there exists a chain $\mathfrak{H}_0 \subset \mathfrak{H}_1 \subset \cdots \subset \mathfrak{H}_r$ of ideals in $\mathfrak{L}$, beginning with $\mathfrak{H}_0 = \{0\}$, ending with $\mathfrak{H}_r = \mathfrak{L}$, such that the factor algebras $\mathfrak{H}_i/\mathfrak{H}_{i-1}$ $(1 \leqq i \leqq r)$ are all abelian. Every sub-algebra or factor-algebra of a solvable Lie algebra is again solvable.

The derived algebra $\mathfrak{L}'$ of a solvable Lie algebra $\mathfrak{L} = \mathfrak{L}_n$ of dimension $n > 0$ is always different from $\mathfrak{L}$. Since $\mathfrak{L}/\mathfrak{L}'$ is abelian, every linear sub-space of $\mathfrak{L}$ containing $\mathfrak{L}'$ is an ideal, and we conclude that $\mathfrak{L}$ has an ideal $\mathfrak{L}_{n-1}$ of dimension $n - 1$. The algebra $\mathfrak{L}_{n-1}$ is also solvable if $n - 1 > 0$; it has itself an ideal $\mathfrak{L}_{n-2}$ of dimension $n - 2$, and we may continue the process. Hence we have a chain $\mathfrak{L}_0 = \{0\} \subset \mathfrak{L}_1 \subset \mathfrak{L}_2 \subset \cdots \subset \mathfrak{L}_n = \mathfrak{L}$ of sub-algebras of $\mathfrak{L}$, where $\mathfrak{L}_i$ is of dimension i and contained as an ideal in $\mathfrak{L}_{i+1}$ (although $\mathfrak{L}_i$ is in general not an ideal in $\mathfrak{L}$). If for every i $(1 \leqq i \leqq n)$ we choose an element L_i belonging to $\mathfrak{L}_i$ but not to $\mathfrak{L}_{i-1}$, the n elements $L_1, L_2, \cdots, L_n$ constitute a base for $\mathfrak{L}$.

Suppose now that $\mathfrak{L}$ is the Lie algebra of a connected simply connected group G. Then every element in G can be written in one and only one way in the form $(\exp t_1L_1)(\exp t_2L_2) \cdots (\exp t_nL_n)$, where $t_1, t_2, \cdots, t_n$ are real numbers.[2] Let $\mathfrak{M}$ be any linear sub-space of $\mathfrak{L}$. We may choose the elements $L_1, L_2, \cdots, L_n$ in such a way as to satisfy the following condition: if there exists in $\mathfrak{L}_i$ an element of $\mathfrak{M}$ which does not belong to $\mathfrak{L}_{i-1}$, L_i shall be chosen in $\mathfrak{M}$. If $L_{i_1} = M_1$, $L_{i_2} = M_2, \cdots, L_{i_r} = M_r$ are the elements L_i which belong to $\mathfrak{M}$, the elements $M_1, M_2, \cdots, M_r$ obviously form a base for $\mathfrak{M}$.

Suppose now that $\mathfrak{M}$ is a sub-algebra. Let k, l be two indices such that $1 \leqq k < l \leqq r$; then $[M_k, M_l]$ is a linear combination of $M_1, M_2, \cdots, M_r$; on the other hand, it is also a linear combination of $L_{i_1}, L_{i_2}, \cdots, L_{i_{l-1}}$; hence it is a linear combination of $M_1, M_2, \cdots, M_{l-1}$. It follows that $M_1, M_2, \cdots, M_l$ constitute a base of a sub-algebra $\mathfrak{M}_l$ of $\mathfrak{M}$ of dimension l, and that $\mathfrak{M}_{l-1}$ is contained as an ideal in $\mathfrak{M}_l$: the base $M_1, M_2, \cdots, M_r$ has with respect to $\mathfrak{M}$ the same properties as the base $L_1, L_2, \cdots, L_n$ with respect to $\mathfrak{L}$.

Suppose that $\mathfrak{L}$ is the Lie algebra of a simply connected group G. Then the elements $(\exp u_1M_1)(\exp u_2M_2) \cdots (\exp u_rM_r)$ $(u_1, u_2, \cdots, u_r$ real) constitute a closed sub-set M of G, homeomorphic to the r-dimensional cartesian space. On the other hand, we know that there exists an abstract Lie group M' whose Lie algebra is $\mathfrak{M}$ and that every element in M' can be written in one and only

[4] Or *integrable* in the sense of S. Lie or E. Cartan.

one way in the form $(\exp u_1'M_1) \cdots (\exp u_r'M_r)$. It follows at once that M is a sub-group of G.

Therefore: *If G is a simply connected solvable Lie group, to every sub-algebra $\mathfrak{M}$ of the Lie algebra of G there corresponds a closed simply connected sub-group M of G whose Lie algebra is $\mathfrak{M}$.*

III. Study of a Particular Type of Solvable Group

Let G be a simply connected solvable Lie group which has the following properties:[5]

1) The Lie algebra $\mathfrak{L}$ of G, of dimension n, has an abelian ideal $\mathfrak{A}$ of dimension $n - 1$.

2) There exists an element σ of the center of G which does not belong to the connected sub-group $\mathfrak{a}$ of G whose Lie algebra is $\mathfrak{A}$.

We shall first prove that there exists an element $L \in \mathfrak{L}$ such that $\exp L = \sigma$. In fact the group G/A is an abelian connected group of dimension 1; hence there exists an element $\bar{L}$ of $\mathfrak{L}/\mathfrak{A}$ such that $\exp \bar{L} = \bar{\sigma}$, where $\bar{\sigma}$ is the coset modulo $\mathfrak{a}$ which contains σ. If L_1 is any element of $\mathfrak{L}$ whose coset modulo $\mathfrak{A}$ is $\bar{L}$, we have $\exp L_1 = \sigma\eta$, $\eta \in \mathfrak{a}$. If t is any real number, $\exp tL_1$ commutes with $\exp L_1$ and with σ, hence also with η. On the other hand, $\mathfrak{a}$ being abelian and connected, there exists one and only one element $A \in \mathfrak{A}$ such that $\eta = \exp A$. In the adjoint representation, the operation which corresponds to $\exp tL_1$ changes A into an element A_t such that $\exp A_t = \eta = \exp A$; hence $A_t = A$, which means that $[AL_1] = 0$; it follows that $\exp (L_1 - A) = \exp L_1 (\exp - A) = \sigma$; we may take $L = L_1 - A$.

Consider now the group $\mathfrak{G}$ composed of the elements $\exp tL$. Since $\exp L$ belongs to the center, the adjoint representation maps $\mathfrak{G}$ onto a group of linear operators in $\mathfrak{L}$ which is compact. The operators of this group permute among themselves the elements of $\mathfrak{A}$; since the group is compact, we have complete reducibility,[6] i.e. we can find a decomposition $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2 + \cdots + \mathfrak{A}_h$ of $\mathfrak{A}$ into a direct sum of linear spaces $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_h$ each one of which gives an irreducible representation of $\mathfrak{G}$. Since $\mathfrak{G}$ is a one dimensional group (isomorphic to the additive group of real numbers), each $\mathfrak{A}_i$ is of dimension 1 or 2; suppose that $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_k$ are of dimension 2, $\mathfrak{A}_{k+1}, \cdots, \mathfrak{A}_h$ being of dimension 1.

The operation A_L which corresponds to L in the adjoint representation permutes among themselves the elements of $\mathfrak{A}_i$; let λ_i be one of its characteristic roots in $\mathfrak{A}_i$. Then e^{λ_i} is a characteristic root of the operation which represents $\exp L = \sigma$; hence $e^{\lambda_i} = 1$, $\lambda_i = 2\pi n_i\sqrt{-1}$, where n_i is an integer. If $n_i \neq 0$, λ_i is not real, and $-\lambda_i$ is also a characteristic root of A_L in $\mathfrak{A}_i$; hence $\dim \mathfrak{A}_i = 2$; if $n_i = 0$ on the contrary, we have $\dim \mathfrak{A}_i = 1$.

[5] The simplest example of such a group is the simply connected covering group of the group of motions in the plane.

[6] Cf. *Pontrjagin*, l. c. 2), Theorem 23, p. 110 and I, p. 109.

Moreover, if $i \leqq k$ we can find a base A_i, B_i of $\mathfrak{A}_i$ such that $\mathrm{A}_L(A_i + \sqrt{-1}\, B_i) = 2\pi n_i \sqrt{-1}\, (A_i + \sqrt{-1}\, B_i)$, i.e.

$$\begin{aligned} [A_i, L] &= -2\pi n_i B_i \\ [B_i, L] &= 2\pi n_i A_i \end{aligned} \qquad (n_i \neq 0).$$

If $i > k$, we have

$$[A_i, L] = 0.$$

Hence the structure of $\mathfrak{L}$ is entirely determined when the integers n_i are known; the linear space $\mathfrak{A}_{k+1} + \cdots + \mathfrak{A}_h$ is the center of $\mathfrak{L}$, and $\mathfrak{A}_1 + \mathfrak{A}_2 + \cdots + \mathfrak{A}_k$ is the derived algebra of $\mathfrak{L}$.

It is easy to build from these data a system of finite equations for the group. We have n variables t, u_i, v_i $(1 \leqq i \leqq k)$, w_i $(k < i \leqq h)$ which constitute a system of coordinates on the group; σ_1, σ_2 being any two elements of the group, the law of composition is defined by the formulas

$$t(\sigma_1, \sigma_2) = t(\sigma_1) + t(\sigma_2) \qquad w_i(\sigma_1\sigma_2) = \exp\left(-2\pi n_i \sqrt{-1}\, t(\sigma_2)\right) \quad (w_i(\sigma_1) + w_i(\sigma_2))$$

where we have set $w_i = u_i + \sqrt{-1}\, v_i$ if $1 \leqq i \leqq k$.[7] The coordinates of σ are $(1, 0, \cdots, 0)$.

We can also find the explicit formulas for the coordinates of $\exp X$, if $X = a_0 L + \sum_{i=1}^{k} (a_i A_i + b_i B_i) + \sum_{k+1}^{h} a_i A_i$: we find

$$t(\exp X) = a_0$$

if $1 \leqq i \leqq k$

$$w_i(\exp X) = \begin{cases} (a_i + \sqrt{-1}\, b_i) \dfrac{1 - \exp\left(-2\pi n_i \sqrt{-1}\, a_0\right)}{2\pi n_i \sqrt{-1}\, a_0} & \text{if } n_i a_0 \neq 0 \\ a_i + \sqrt{-1}\, b_i & \text{if } n_i a_0 = 0 \end{cases}$$

if $i > k$ $\qquad w_i(\exp X) = a_i.$

Hence, the elements X for which the equality $\exp X = \sigma$ holds are those of the form $L + \sum_1^k (a_i A_i + b_i B_i)$, and only those.

We shall need later the following consequence of these facts: let τ be any automorphism of the group G which leaves unchanged the element σ; if we denote by A_τ the linear operation[8] on $\mathfrak{L}$ which corresponds to τ, A_τ permutes among themselves the elements of the sub-algebra $\mathfrak{M}$ generated by L and the elements A_i, B_i for $1 \leqq i \leqq k$; if X is any element of $\mathfrak{M}$ such that $X \equiv L \pmod{\mathfrak{A}}$, we have again $\exp X = \sigma$.

[7] In order to simplify the writing of formulas, we are using complex valued functions w_i, although the group under consideration is of course a *real* Lie group.

[8] This operation is defined just as in the case of the adjoint-representation.

IV. The Central Theorem

We want to prove the following result:

Theorem 1. *Let G be a connected simply connected solvable group, with the Lie algebra $\mathfrak{L}$. Let D be a discrete sub-group of the center of G. Then D is a free abelian group of a certain rank r at most equal to the dimension n of G. It is possible to choose a base $L_1, L_2, \cdots, L_n$ of $\mathfrak{L}$ which has the following properties:*

1) *Every element in G can be written in one and only one way in the form* $(\exp t_1L_1)(\exp t_2L_2) \cdots (\exp t_nL_n)$.

2) *There exists a subset $\{i_1, i_2, \cdots, i_r\}$ of the set $\{1, 2, \cdots, n\}$ such that the elements* $\exp L_{i_1}, \exp L_{i_2}, \cdots, \exp L_{i_r}$ *constitute a base for D, and that $[L_{i_\alpha}, L_{i_\beta}] = 0 (1 \leqq \alpha, \beta \leqq r)$.*

We shall prove this theorem by induction on the dimension n of $\mathfrak{L}$. Let us assume that it has been proved already for all dimensions $< n$.

Since $\mathfrak{L}$ is solvable, it has certainly an abelian ideal of dimension > 0; let $\mathfrak{A}$ be an abelian ideal of maximal dimension; we set $\mathfrak{L}^* = \mathfrak{L}/\mathfrak{A}$. Let A be the connected abelian sub-group of G whose Lie algebra is $\mathfrak{A}$; A is an invariant sub-group, and G/A is a solvable group of dimension $< n$. The group DA/A is a sub-group of the center of G/A.

Consider now the closure $\overline{DA}$ of the group DA. The connected component of the unit element ϵ in this closure is a closed sub-group of G; according to a well known theorem of E. Cartan, it is a Lie group; let $\mathfrak{A}'$ be its Lie algebra. The group $\overline{DA}$ being abelian and invariant, $\mathfrak{A}'$ is an abelian ideal in $\mathfrak{L}$ containing $\mathfrak{A}$; in virtue of our choice of $\mathfrak{A}$, we have $\mathfrak{A}' = \mathfrak{A}$, which shows that the connected component of ϵ in $\overline{DA}$ reduces to A. On the other hand, there is a neighborhood of ϵ in $\overline{DA}$ which is compact; this neighborhood can meet only a finite number of distinct connected components of $\overline{DA}$, from which it follows at once that there exists a neighborhood U of ϵ in G such that $U \cap \overline{DA} \subset A$. Any element of DA which belongs to the set UA already belongs to A; hence the group DA/A is a discrete sub-group of G/A. It follows that we can apply our theorem to G/A and DA/A: we obtain a base $\{L_1^*, L_2^*, \cdots, L_m^*\}$ of $\mathfrak{L}/\mathfrak{A}$ with the following properties:

A) Every element of G/A can be written in one and only one way in the form $(\exp t_1L_1^*)(\exp t_2L_2^*) \cdots (\exp t_mL_m^*)$.

B) There exists a subset $\{i_1, i_2, \cdots, i_s\}$ of the set $\{1, 2, \cdots, m\}$ such that the elements exp. $L_{i_1}^*$, $\exp L_{i_2}^*, \cdots, \exp L_{i_s}^*$ constitute a base for the free discrete group DA/A, and that $[L_{i_\alpha}^*, L_{i_\beta}^*] = 0$ $(1 \leqq \alpha, \beta \leqq s)$.

For each α, we can find an element $\sigma_\alpha \in D$ such that its coset σ_α^* modulo A coincides with $\exp L_{i_\alpha}^*$.

We want now to choose in a certain way elements $L_1, L_2, \cdots, L_m$ of $\mathfrak{L}$ which will have to satisfy the following conditions: A) The coset of L_i modulo $\mathfrak{A}$ is L_i^* $(1 \leqq i \leqq m)$; B) we have $\exp L_{i_\alpha} = \sigma_{i_\alpha}$ $(1 \leqq \alpha \leqq s)$; C) we have $[L_{i_\alpha}, L_{i_\beta}] = 0$ $(1 \leqq \alpha, \beta \leqq s)$.

Suppose that we have already chosen $L_1, L_2, \cdots, L_h$ in such a way as to

satisfy the conditions of the type A), B), C) which concern these elements. We want to choose L_{h+1} (If $h = 0$, we assume simply that no choice has been made yet).

If $h + 1$ does not occur among the indices $i_1, i_2, \cdots, i_s$ we choose for L_{h+1} any element of $\mathfrak{L}$ whose coset modulo $\mathfrak{A}$ is L^*_{h+1}.

Suppose now that $h + 1 = i_{t+1}$. We begin by choosing an L'_{h+1} whose coset modulo $\mathfrak{A}$ is L^*_{h+1}, and we consider the Lie algebra $\mathfrak{H}$ spanned by $\mathfrak{A}$ and L'_{h+1}; let H be the corresponding connected subgroup of G. It is a closed simply connected sub-group of G. Moreover $\exp L'_{h+1}$ lies in the coset $\exp L^*_{h+1}$ modulo A; hence $\exp L'_{h+1} \equiv \sigma_{i_{t+1}}$ modulo A, which proves that $\sigma_{i_{t+1}} \in H$; since $\sigma_{i_{t+1}} \notin A$, H is a group of the kind which has been considered in §III. It follows first that $\mathfrak{H}$ contains an element L''_{h+1} such that $L''_{h+1} \equiv L'_{h+1}$ (modulo $\mathfrak{A}$) and that $\exp L''_{h+1} = \sigma_{i_{t+1}}$. Since we have $[L^*_{i_\alpha}, L^*_{h+1}] = 0$ $(1 \leqq \alpha \leqq s)$, we have $[L_{i_\gamma}, L''_{h+1}] \in \mathfrak{A}$ $(1 \leqq \gamma \leqq t)$. Since $[L_{i_\gamma}, L_{i_{\gamma'}}] = 0$ for $1 \leqq \gamma,\gamma' \leqq t$, the elements $(\exp u_1 L_{i_1}) \cdots (\exp u_t L_{i_t})$ constitute an abelian sub-group of G. Since $\mathfrak{A}$ is an ideal and $[L_{i_\gamma}, L''_{h+1}] \in \mathfrak{A}$, they are represented in the adjoint representation by linear operations which permute among themselves the elements of $\mathfrak{H}$ and $\mathfrak{A}$. The elements $\exp L_{i_\gamma}$ $(1 \leqq \gamma \leqq t)$ being represented by the unit operation, we obtain in this way a compact abelian group $\mathfrak{G}$ of linear transformations of $\mathfrak{H}$ into itself.

It follows from §III that $\mathfrak{H}$ can be represented in the form $\mathfrak{H}_1 \times \mathfrak{Z}$, where $\mathfrak{Z}$ is in the center of $\mathfrak{H}$ and $L''_{h+1} \in \mathfrak{H}_1$. Moreover the decomposition has, as we saw, the two further properties that: 1) if $X \in \mathfrak{H}_1$, $X \equiv L''_{h+1} (\mathrm{mod}\ \mathfrak{A})$, we have $\exp X = \exp L''_{h+1} = \sigma_{i_{t+1}}$ 2) any automorphism of H which leaves $\sigma_{i_{t+1}}$ invariant gives a linear operation in $\mathfrak{H}$ which permutes among themselves the elements of $\mathfrak{H}_1$.

This applies in particular to the inner automorphisms of G produced by the elements $(\exp u_1 L_{i_1})(\exp u_2 L_{i_2}) \cdots (\exp u_t L_{i_t})$. Hence the operations of the group $\mathfrak{G}$ permute among themselves the elements of $\mathfrak{H}_1$.

But $\mathfrak{H}_1 \cap \mathfrak{A}$ is a linear sub-space of $\mathfrak{H}_1$ whose dimension is exactly one less than the dimension of $\mathfrak{H}_1$. Its elements are also permuted among themselves by the operations of $\mathfrak{G}$. Since $\mathfrak{G}$ is compact, it is completely reducible, and we may conclude that there exists an element $L_{h+1} \in \mathfrak{H}_1$ with the following properties: 1) $L_{h+1} \equiv L''_{h+1}$ (mod $\mathfrak{A}$); 2) the elements of $\mathfrak{G}$ permute among themselves the multiples tL_{h+1} of L_{h+1}. Again because $\mathfrak{G}$ is compact, it follows that its operations leave L_{h+1} invariant; if we remember the definition of $\mathfrak{G}$, we see that $[L_{i_\gamma}, L_{h+1}] = 0$ $(1 \leqq \gamma \leqq t)$. Finally, the conditions $L_{h+1} \in \mathfrak{H}_1$, $L_{h+1} \equiv L''_{h+1}$ (mod $\mathfrak{A}$) imply $\exp L_{h+1} = \exp L''_{h+1} = \sigma_{i_{t+1}}$. Hence the element L_{h+1} satisfies all our requirements.

In this way, we shall be able to select m elements $L_1, L_2, \cdots, L_m$ satisfying the conditions A, B, C.

On the other hand, the group A is isomorphic to the additive group of R^{n-m}, and $D \cap A$ is a discrete sub-group of A, i.e. a lattice. It follows that we can find elements $L_{m+1}, \cdots, L_n$ in $\mathfrak{A}$ such that every element of A can be written

in one and only one way in the form $(\exp t_{m+1}L_{m+1}) \cdots (\exp t_nL_n)$ and that there exists a subset $\{i_{s+1}, \cdots, i_r\}$ of the set $\{m+1, \cdots, n\}$ such that the elements $\exp L_{i_{s+1}}, \cdots, \exp L_{i_r}$ constitute a base for $D \cap A$.

It remains only to prove that the sets $\{L_1, L_2, \cdots, L_n\}$ and $\{i_1, i_2, \cdots, i_r\}$ which have been selected have the properties 1), 2) of our theorem.

Let σ be any element of G and σ^* be its coset modulo A. Then we can write σ^* in the form $(\exp t_1L_1^*) \cdots (\exp t_mL_m^*)$. But this product represents a coset modulo A which contains the element $(\exp t_1L_1)(\exp t_2L_2) \cdots (\exp t_mL_m)$ since L_m belongs to the coset L_m^* modulo $\mathfrak{A}$. Hence we have

$$(1) \qquad \sigma = (\exp t_1L_1)(\exp t_2L_2) \cdots (\exp t_mL_m)\zeta \qquad \zeta \in A$$

and ζ can be written in the form $(\exp t_{m+1}L_{m+1}) \cdots (\exp t_nL_n)$, which proves that σ is equal to an element of the form $(\exp t_1L_1)(\exp t_2L_2) \cdots (\exp t_nL_n)$. Moreover $t_1, t_2, \cdots, t_m$ are uniquely determined by σ^* (and *a fortiori* by σ); hence ζ is determined by σ and also $t_{m+1}, \cdots, t_n$.

If $\sigma \in D$, we have $\sigma^* \in DA/A$ and therefore σ^* can be written in the form $(\exp m_{i_1}L_{i_1}^*) \cdots (\exp m_{i_s}L_{i_s}^*)$, with integral $m_{i_1}, \cdots, m_{i_s}$. Since the element $(\exp m_{i_1}L_{i_1}) \cdots (\exp m_{i_s}L_{i_s}) = \sigma_{i_1}^{m_{i_1}}\sigma_{i_2}^{m_{i_2}} \cdots \sigma_{i_s}^{m_{i_s}}$ belongs to D, the element ζ of formula (1) belongs to D and can be written in the form $(\exp m_{i_{s+1}}L_{i_{s+1}}) \cdots (\exp m_{i_r}L_{i_r})$, which proves that the elements $\exp L_{i_1}, \exp L_{i_2}, \cdots, \exp L_{i_r}$ constitute a base for D.

We have already seen that $[L_{i_\alpha}, L_{i_\beta}] = 0$ for $1 \leqq \alpha, \beta \leqq s$. We shall see that, for $\alpha > s$, L_{i_α} necessarily belongs to the center of $\mathfrak{L}$. In fact, let us consider an arbitrary operation of the adjoint representation of G; it changes L_{i_α} in an element L'_{i_α} which still belongs to $\mathfrak{A}$; since $\sigma_\alpha = \exp L_{i_\alpha}$ belongs to the center we have $\exp L'_{i_\alpha} = \exp L_{i_\alpha}$. Since A is an abelian simply connected group, we conclude that $L'_{i_\alpha} = L_{i_\alpha}$, which proves that L_{i_α} belongs to the center of $\mathfrak{L}$. This completes the proof of our central theorem.

Remark. We may assume that the set $\{i_1, i_2, \cdots, i_r\}$ is the set $\{1, 2, \cdots, r\}$. We again prove it by induction on n. Suppose that we have $\{i_1, i_2, \cdots, i_s\} = \{1, 2, \cdots, s\}$. We have seen that $L_{i_{s+1}}, \cdots, L_{i_r}$ belong to the center of $\mathfrak{L}$. It follows that any element $\exp t_{i_\alpha}L_{i_\alpha}$ $(\alpha > s)$ belongs to the center of G. Hence the property 1 stated in the theorem will still hold if we operate a permutation of $L_1, L_2, \cdots, L_n$ which brings L_{i_α} at α-th place $(\alpha = s+1, \cdots, r)$, and this proves our assertion.

Theorem 9. *Let G^* be a connected solvable Lie group. There exists a compact abelian sub-group T of G^* and a sub-set E of G^*, homeomorphic to a cartesian space in a certain number of dimensions, such that every element $\sigma \in G^*$ can be written in one and only one way in the form $\sigma = \tau\eta$, $\tau \in T$, $\eta \in E$. The mapping $(\tau, \eta) \rightarrow \tau\eta$ is a homeomorphism of $T \times E$ with G^*.*

In fact we have $G^* = G/D$ where G is a simply connected solvable Lie group and D a discrete sub-group of the center of G. We can find a base $\{L_1, L_2, \cdots, L_n\}$ of the Lie algebra of G such that every element in G can be written in one and only one way in the form $(\exp t_1L_1) \cdots (\exp t_nL_n)$, and that

the elements $\exp L_1$, $\exp L_2$, $\cdots$, $\exp L_r$ constitute a base for D. Moreover the group A of the elements $(\exp t_1L_1) \cdots (\exp t_rL_r)$ is abelian.

We have $D \subset A$ and A/D is a compact abelian sub-group T of G^*.

The elements $(\exp t_1L_1) \cdots (\exp t_nL_n)$ and $(\exp t_1'L_1) \cdots (\exp t_n'L_n)$ are congruent modulo D if and only if $t_1' \equiv t_1 \pmod 1$, $\cdots$, $t_r' \equiv t_r \pmod 1$, $t_{r+1}' = t_{r+1}$, $\cdots$, $t_n' = t_n$. Let φ be the natural mapping of G onto G^*. The set $\tilde{E}$ of the elements of the form $(\exp t_{r+1}L_{r+1}) \cdots (\exp t_nL_n)$ is mapped by φ in one-to-one continuous way onto a subset E of G^*. We claim that this mapping is a homeomorphism. In fact, let $\tilde{U}$ be the set of the elements $(\exp t_1L_1) \cdots (\exp t_nL_n)$ with $| t_1 | < \frac{1}{2}$, $\cdots$, $| t_r | < \frac{1}{2}$; $\tilde{U}$ is a neighborhood of $\tilde{E}$ in G and is mapped in a one-to-one continuous way by φ; but we know that φ maps any open subset of G onto an open subset of G^*; hence φ gives a homeomorphism of U with a subset of G^*, which proves our assertion.

It follows that E is homeomorphic to the $(n - r)$-dimensional cartesian space. Moreover every element of G^* can obviously in one and only one way be written in the form $\tau\eta$, $\tau \in T$, $\eta \in E$.

Since the mapping $(\tau, \eta) \rightarrow \tau\eta$ clearly maps open sets of $T \times E$ onto open sets of G^*, it is a homeomorphism, and theorem 2 is completely proved.

Since a compact abelian group is a torus-space (product of circumferences) we have the

THEOREM 2^a. *Any connected solvable Lie group is homeomorphic to the product of a torus space and of a cartesian space.*

PRINCETON UNIVERSITY

Annals of Mathematics
Vol. 42, No. 3, July, 1941

ALTERNATIVE RINGS AND RELATED QUESTIONS I: EXISTENCE OF THE RADICAL

By Max Zorn

(Received July 29, 1940)

In a previous publication (1) I have presented a theory of alternative rings which led up to the characterization of semisimple systems. The present paper is concerned with general hypercomplex[1] alternative rings; its main object is to establish the existence of the radical. In other words it is shown that the properly nilpotent elements form a (two-sided) ideal.

The first section contains the proof of several alternative identities,[2] which are then applied (sec. 2) to the theory of the inverse element. Though a certain amount of duplication[3] could not be avoided, the new arrangement has been given in detail because it is phrased almost entirely in terms of "associators" $(a, b, c) = (ab)c - a(bc)$. A new proof of the O-lemma[4] has likewise been included; the lèmma is now obtained by inspection of an identity, and the assumption about characteristics is removed. This is of interest in connection with the theory of Peirce decompositions, which was derived in (1) under the (now unnecessary) restriction that the characteristics be different from 2.

Thus far the theory is entirely self-contained; this is likewise achieved in the third section by formulating two theorems from (1) as assumptions. These two propositions are indeed so simple that they may very well serve as axioms. On this basis it is possible to treat the rings which in the associative theory have been called "vollstaendig primaer";[5] the existence of the radical is verified in this important special case.

In the last two sections use is made of the theory of idempotents and Peirce decompositions as developed in (1); the necessary material is very briefly reviewed in sec. 4. A certain ideal N is constructed in the fifth section, which depends formally on an arbitrary but fixed maximal Peirce decomposition. The concluding section contains the proof that this ideal consists exactly of all properly nilpotent elements and may therefore be regarded as the radical of the ring.

Notations. With one exception (the delta in the Kronecker symbol δ_{ij}) we use only italics; i, j, k, l, m, n for indices and the like, $a, b, c, d, e, x, y_{ik}, z_{ii}$ for elements, capitals $S, S_{ik}, N_{ik}, N, A, B$ for sets.

[1] See sec. 4 and also (1) p. 137.

[2] I am indebted to P. Jordan for many remarks concerning such and more general identities.

[3] See (1) p. 142.

[4] See (1) pp. 129, 130.

[5] See (2) p. 256 and (3) p. 85.

The summation sign is used in such a way that the range of the index is described in a parenthesis which precedes the letter $\sum$.

Theorems and definitions are numbered decimally within the section and the number of the section is prefixed; a theorem is indicated by a parenthesis—(3.7.1)—and a definition by a bracket—[1.1]—. A single numeral in a parenthesis, like (2), refers to the bibliography.

1. We shall constantly have to use the abbreviation:

$$[1.1] \qquad (a, b, c) = (ab)c - a(bc).$$

These "associators" satisfy in every distributive ring the identity

$$(1.2) \qquad (ab, c, d) - (a, bc, d) + (a, b, cd) = a(b, c, d) + (a, b, c)d$$

which becomes trivial when the associators are worked out according to [1.1].

Let us permute the arguments cyclically such that b is the first letter, and add the resulting equation to (1.2); we get

$$(1.3) \quad (ab, c, d) + (b, c, da) + [(bc, d, a) - (a, bc, d)] + [(a, b, cd) - (b, cd, a)]$$
$$= a(b, c, d) + (a, b, c)d + b(c, d, a) + (b, c, d)a.$$

From now on we ask that multiplication in the distributive ring satisfy the alternative law, which can be expressed in different ways:

[1.4] A ring is called *alternative* and said to satisfy the alternative law of multiplication if the associator (a, b, c) is an alternating function of its arguments. The associator must therefore be zero if two of the arguments are equal; it takes on the factor -1 if two of the arguments are transposed and it remains unchanged if the arguments undergo a cyclic permutation:

$$(1.4.1) \quad (a, a, b) = (a, b, a) = (b, a, a) = 0,$$
$$(a, b, c) + (b, a, c) = (a, b, c) + (a, c, b) = (a, b, c) + (c, b, a) = 0,$$
$$(a, b, c) = (b, c, a) = (c, a, b).$$

The left hand member of the equation (1.3) would consequently reduce to $(ab, c, d) + (b, c, da)$; if we put $b = c$, most of the terms will drop out and the equation $(ab, b, d) = b(b, d, a)$ remains. Exchanging a and b and applying the alternative law we get the identity

$$(1.6) \qquad (a, ba, c) = -a(b, a, c).^{6}$$

We add this equation to

$$(ab, a, c) - (a, ba, c) + (a, b, ac) = a(b, a, c) + (a, b, a)c$$

[6] (1) p. 142.

such as to get

$$(1.7) \qquad (ab, a, c) + (a, b, ac) = 0.$$

On this identity we shall base, in the following section, the theory of the inverse element.

From (1.7) we pass to the next identity

$$(1.8) \qquad (ab, a, c) + (ba, a, c) - (b, a^2, c) = 0$$

by way of adding

$$(ba, a, c) - (b, a^2, c) + (b, a, ac) = b(a, a, c) + (b, a, a)c = 0.$$

2. The theory of Peirce decompositions, which was given in (1), rests on the "O-lemma":

(2.1) *If $ab = ba = 0$ then for every c the associator (a^2, b, c) vanishes.*

Without any restriction this lemma follows by substitution from the identity (1.8); thus we are entitled to apply all results of (1) about idempotent elements without reservation. This will be of importance in the concluding sections.

For the remainder of this section we deal with the theory of the inverse and assume consequently the existence of a unit element.

Once more we apply the identity (1.8), this time for $ab = ba = 1$:

(2.2) *If $ab = ba = 1$, then all associators of the forms (a^2, b, c) and (a, b^2, c) are zero.*

In order to extend this to associators of the type (a, b, c) we replace b by b^2 in the formula (1.7):

$$(2.3) \qquad (ab^2, a, c) + (a, b^2, ac) = 0.$$

Since in accordance with the alternative law the equation $ab = 1$ implies that $ab^2 = a(bb) = (ab)b = 1b = b$, we get from (2.2) and (2.3) the theorem:

(2.4) *If the numbers a and b satisfy the equations $ab = ba = 1$, then every associator of the form (a, b, c) vanishes.*[7]

The element b (if it exists) has all properties of an inverse element for a; it is unique and may be used for the solution of linear equations:

(2.5) *If $ab = ba = 1$, then the equations $ax = c$ and $ya = c$ have the solutions $x = bc$ and $y = cb$; these solutions are unique.*

Indeed, the preceding theorem permits us to proceed as is customary for associative systems. The existence of the solutions follows from formulas like $a(bc) = (ab)c - (a, b, c) = 1c - 0 = c$, and the uniqueness is exhibited by similar equations like $cb = (ya)b = y(ab) + (y, a, b) = y1 + 0 = y$.

In particular, by considering the cases $c = 1$ and $c = 0$, we obtain the corollary:[8]

(2.6) *If $ab = ba = 1$, then b is the unique solution of either one of the two equations $ax = 1$ or $ya = 1$; the element a is not a divisor of zero.*

[7] Compare (1) p. 142, lemma 2.

[8] For additional information about the inverse in alternative fields see (4).

This justifies the definition:

[2.7] If there exists an element b satisfying $ab = ba = 1$, then it is called the *inverse* of a and denoted by a^{-1}.

This definition and the foregoing results permit us to state:

$$(2.8)\qquad aa^{-1} = a^{-1}a = 1; \qquad (a^{-1})^{-1} = a; \qquad (a, a^{-1}, c) = 0.$$

3. In order to postpone as far as possible any reference to outside material we introduce the following two assumptions.

(3.1) *Any two elements a, b of an alternative ring generate, by multiplications and additions, an associative subring.*

That this assumption is satisfied in every alternative ring has been shown by Artin;[9] one particular consequence is that the powers a^n with positive exponents are commutative and associative with each other.

(3.2) *An element a is either nilpotent, or else there exists a right multiple $e = ab$ which is idempotent, such that $e^2 = e \neq 0$.*[10]

[3.3] Definition. An alternative ring will be called *primitive* if (a) it contains a unit element 1 which is different from 0; (b) the unit element is the only idempotent of the ring; (c) it satisfies the two above assumptions (3.1) and (3.2).

In defining next the properly nilpotent elements we deviate from the classical definition[11] by including 0. This is done because otherwise we would always have to say "properly nilpotent or zero."

[3.4] An element a is called *properly nilpotent* if all multiples ax are nilpotent. By virtue of (3.1) we see that $(ax)^m = 0$ implies $(xa)^{m+1} = x(ax)^m a = 0$, such that also the left multiples of a are nilpotent; we shall have to use this remark later.

(3.5) *If a multiple ax has an inverse then the element a is not nilpotent.*[12]

Let $a^{m-1} \neq 0 = a^m$; then $0 \neq a^{m-1} = a^{m-1}((ax)(ax)^{-1}) = (a^{m-1}(ax))(ax)^{-1} = (a^m x)(ax)^{-1}$ shows that $a^m = 0$ is impossible.

The theorems (3.6)–(3.9) describe some important properties of primitive rings.

(3.6) *If an element a of a primitive ring is not nilpotent then it has an inverse.*

Indeed there must be a multiple ab which is idempotent, by virtue of (3.2). This idempotent must be equal to 1 since our ring is primitive; the theorem will be proved if we show that $ba = 1$.

First of all we see that ba is not equal to zero; for in that case $1 = 1^2 = (ab)(ab) = a(ba)b = 0$ would ensue.[13] Secondly ba is idempotent because of $(ba)(ba) = b(ab)a = (b1)a = ba$; consequently we have $ba = 1$ and b is the inverse of a.

[9] See (1) p. 127.

[10] This assumption appears in (1) p. 138 as a theorem, valid in hypercomplex rings and incidentally based on the first and a part of the second chain condition.

[11] See (3) p. 46.

[12] Compare (1) p. 142.

[13] This part was omitted in (1) p. 142.

(3.7) *Every nilpotent element of a primitive ring is properly nilpotent.*

(3.7.1) Corollary. *Every element of a primitive ring is either properly nilpotent or it has an inverse.*

In other words if a is nilpotent then every multiple is nilpotent. That is an immediate consequence of (3.5) and (3.6). For if—to conclude indirectly—one multiple ax were not nilpotent it would possess an inverse and (3.5) imply that a is not nilpotent. Using this and the remark about [3.4] we find:

(3.7.2) *If an element a of a primitive ring is properly nilpotent, every multiple ax and xa is properly nilpotent.*

Thus one of the two essential properties of two-sided ideals is verified for the set of all properly nilpotent elements. It remains to show that this set is a modulus (i.e. closed under addition and subtraction). This is accomplished if we prove the theorem:

(3.8) *If a and b are two properly nilpotent elements of a primitive ring, then the the difference $a - b$ is likewise properly nilpotent.*

An equivalent proposition would be:

(3.8.1) *If the element b of a primitive ring is properly nilpotent, and the difference $a - b$ is not, then the element a is not (properly) nilpotent.*

For in that case the difference $a - b$ has in inverse; let $(a - b)x = ax - bx = 1$, so that $ax = 1 + bx$. Since bx is nilpotent, with say $(bx)^m = 0$, the element $1 + bx$ has the inverse $(i = 0 \cdots m) \sum (-bx)^i$. Thus the element a is seen to have a multiple with an inverse; and that precludes, according to the theorem (3.5), the possibility that a be nilpotent.

If we observe finally that the set of all properly nilpotent elements is not empty we may enunciate the theorem:

(3.9) *The properly nilpotent elements of a primitive ring form a two-sided ideal.*

We call this ideal the radical of the primitive ring and define more generally:

[3.10] If the properly nilpotent elements of an alternative ring form an ideal, then this ideal is called the *radical* of the ring in question.

We note without proof:

(3.11) *The residue class ring modulo the radical (if it exists) is semisimple;*

(3.12) *The residue class ring modulo the radical of a primitive ring is an alternative field;*

and

(3.12.1) *If a primitive ring does not contain properly nilpotent elements different from zero, or in other words if the zero ideal is the radical, then the ring is an alternative field.*

We intend to establish the existence of a radical for all hypercomplex rings; for that reason we state as an immediate consequence of (3.2) and [3.10] the fact:

(3.13) *If a ring does not contain any idempotent elements, it possesses a radical, which coincides with the ring itself.*

4. In order to complete our proof we shall have to use some of the definitions and theorems which were developed in (1).

[4.1.1] The set of all right multiples ax of an element a (*I-modulus*) is denoted by aS;

[4.1.2] The solutions of $ax = 0$ form a "*II-modulus*" denoted by $(a\backslash 0)$;

[4.1.3] The modulus generated by the products ab where a and b are respectively taken from the moduli A and B is denoted by AB.

[4.2] An alternative ring is called *hypercomplex* if it satisfies the following three chain conditions:

(4.2.1) every (necessarily descending) chain $a^i S$ is finite;

(4.2.2) every ascending chain of II-moduli $(a^i\backslash 0)$ is finite;

(4.2.3) every descending chain of II-moduli $(a_i\backslash 0)$ is finite.

For such rings the following theorems were derived:[14]

(4.3) *There exists (at least) one finite set $e_1 \cdots e_n$ of idempotents with $e_i e_k = 0$ whenever i and k are different, such that the ring S is the direct sum*

$$S = (i, k = 0 \cdots n) \sum S_{ik}$$

of moduli S_{ik}, the modulus S_{ik} consisting exactly of the elements x_{ik} which satisfy

$$e_j x_{ik} = \delta_{ji} x_{ik}, \qquad x_{ik} e_j = \delta_{kj} x_{ik};$$

the "component" of x in the modulus S_{ik} is denoted by x_{ik} and for $i \neq 0 \neq k$ is given by $e_i x e_k$; this "Peirce-decomposition" has the following properties (4.3.1) *and* (4.3.2).

(4.3.1) *The moduli S_{0i} and S_{i0}, for $i = 0 \cdots n$, consist of properly nilpotent elements; the modulus S_{00} is a subring.*

(4.3.2) *The moduli S_{ii}, for $i = 1 \cdots n$, are hypercomplex and primitive subrings; their respective unit elements are the idempotents e_i.*

The last two properties express the fact that the decomposition is a maximal one. Without loss of generality[15] we assume that there are actually idempotent elements, and that consequently the number n is not 0. Moreover it can be shown by a refinement of the construction in (1) that an additional condition can be satisfied, to wit

(4.4) *If e is an idempotent element, then the Peirce decomposition may be chosen in such a way that e is a sum of elements e_i, or*

$$e = (i = 1 \cdots m) \sum e_i .$$

The next theorem deals with the products of the moduli S_{ik}. Like the remaining statements from this section it is valid for general Peirce decompositions.

(4.5.1) $$S_{ik} S_{kl} \subset S_{il},$$

(4.5.2) $$S_{ik} S_{ik} \subset S_{ki},$$

[14] Under the restriction that $a + a = 0$ imply $a = 0$; we have removed this restriction in the second section.

[15] See (3.13) and (3.2)

(4.5.3) All other products $S_{ik}S_{lm}$ are zero.

The second of these statements is amplified by the formula:

(4.6.1) $$x_{ik}y_{ik} + y_{ik}x_{ik} = 0 \quad if \quad i \neq k;$$

this fact is going to be used in the weaker form:

(4.6.2) *If N_{ik} is a modulus contained in S_{ik}, and if $i \neq k$, then we can say that $N_{ik}S_{ik} = S_{ik}N_{ik}$.*

Concerning products of three factors only a part of the "Dreierregeln"[16] is necessary for our purposes, which we report in the next theorem:

(4.7.1)[17] *With the exception of products like $(x_{ii}y_{ii})z_{ii}$ every product of type $(x_{ik}y_{kl})z_{li}$ is associative, that is, equal to $x_{ik}(y_{kl}z_{li})$.*

(4.7.2)[18] *A product of type $(x_{ik}y_{ik})z_{ik}$, where the indices i and k are different retains its value if the factors undergo a cyclic permutation.*

We name finally the rules[19] about powers of the form $(a_{ii}x)^m$ and $(a_{ik}x)^m$:

(4.8.1) $$(a_{ii}x)^m = (a_{ii}x_{ii})^{m-1}(j = 0 \cdots n) \sum a_{ii}x_{ij} ;$$

and for $i \neq k$

(4.8.2) $$(a_{ik}x)^m = (a_{ik}x_{ki})^{m-1}(j = 0 \cdots n) \sum a_{ik}x_{kj} + (a_{ik}x_{ik})(a_{ik}x_{ki})^{m-1}.$$

If we compare, in the latter formula, the components in S_{ii} of the two members, we see that

(4.8.3) $e_i(a_{ik}x)^m e_i = (a_{ik}x_{ki})^m$, *and this holds likewise if the two indices are equal.*

The formulas (4.8) will only be used in the last section.

5. With the aid of theorems developed in 4 and reported about in 5 we shall now, for a fixed but arbitrary Peirce decomposition with the properties enumerated in section 4, construct an ideal which later on will turn out to be the

---

[16] See (1) p. 134.

[17] See (1) p. 136, last paragraph of sec. 3.

In order to obtain the theorem (4.7.1) from (1) it is necessary to distinguish various cases; the following direct proof seems more satisfactory. We see that if two of the ordered index pairs ij, jk, ki are equal, then the three indices are equal; for due to the cyclic symmetry of the situation we may assume in such a case that the first two pairs are identical, and that would mean $i = j$ and $j = k$. In those cases which are not classed as exceptions by the wording of the theorem we have to deal with three different pairs of indices, and we may consider the elements in question as components of a single element, namely, the sum $x_{ij} + y_{jk} + z_{ki} = a$. But the components of one element a with respect to a fixed Peirce decomposition are contained in the subring generated by a and the idempotents e_l.

A single element generates an A-ring (see (1) p. 127); and so does, by virtue of the O-lemma, and the A-lemma, any set of orthogonal elements. According to the generalized theorem of Artin two A-rings generate an associative ring, and thus the components of a single number are seen to be associative with each other. In particular the three numbers x_{ij}, y_{jk}, and z_{ki} are associative, q.e.d.

[18] See in (1) on p. 135 the more general formula (23).

[19] See (1) p. 136.

radical of the ring S. This ideal N will consist exactly of the quantities whose components are "singular" according to the following definition.

[5.1] An element a_{ik} (of S_{ik}) is called *singular* if all elements of $a_{ik}S_{ki}$ are nilpotent, otherwise regular.[20]

The following statements are easily deduced from the theorems which are indicated thereafter.

(5.2.1) *If one of the indices i or k is zero then every element a_{ik} is singular.* See (4.3.1).

(5.2.2) *For equal indices the elements a_{ii} constitute the radical of the ring S_{ii}.* See (4.3.2) and (3.13).

(5.2.3) *If a_{ik} is singular, all elements of $S_{ki}a_{ik}$ are likewise nilpotent.* See the remark about [3.4].

We define:

[5.3] The set of all singular elements in S_{ik} shall be denoted by N_{ik}.

We see that the sets N_{i0} and N_{0i} are identical with S_{i0} and S_{0i} respectively and are therefore moduli; likewise—according to (5.2.2)—the sets N_{ii} are the radicals of S_{ii} and are therefore moduli. We maintain

(5.4) *The sets N_{ik} are moduli in any case.*

Suppose that both elements a_{ik} and b_{ik} are singular, and consider the difference $a_{ik} - b_{ik}$. Every multiple $(a_{ik} - b_{ik})x_{ki}$ is the difference of two numbers, namely, $a_{ik}x_{ki}$ and $b_{ik}x_{ki}$, from the radical of S_{ii}; such differences are necessarily in the radical N_{ii}.

The definition of the sets N_{ik} implies that $N_{ik}S_{ki} \subset N_{ii}$, and from (5.2.3) we infer also that $S_{ik}N_{ki} \subset N_{ii}$. Again we make a statement which is slightly more general:

(5.5) *The products $N_{ij}S_{jk}$, $S_{ij}N_{jk}$, $N_{ki}S_{ki}$ and $S_{ki}N_{ki}$ are contained in the modulus N_{ik}.*

For i equal to k nothing new is asserted; for the other cases it is necessary and sufficient to verify the following inclusion relations:

(5.5 1) $$(N_{ij}S_{jk})S_{ki} \subset N_{ii},$$

(5.5.2) $$(S_{ij}N_{jk})S_{ki} \subset N_{ii},$$

(5.5.3) $$(N_{ki}S_{ki})S_{ki} \subset N_{ii},$$

(5.5.4) $$(S_{ki}N_{ki})S_{ki} \subset N_{ii}.$$

Evidently the last of these formulas is equivalent to the third, on account of (4.6.2). According to (4.7.2) the third relation is equivalent to $(S_{ki}S_{ki})N_{ki} \subset N_{ii}$, *and this would follow* from the second formula because we know that $S_{ik}S_{ik} \subset S_{ki}$. The second inclusion in turn would be a consequence of the first. For a product of this type is associative,[21] therefore instead of (5.5.2) the rela-

[20] Compare (2) p. 257.

[21] See (4.7.1).

tion of $S_{ij}(N_{jk}S_{ki}) \subset N_{ii}$ may be discussed; and if indeed $N_{jk}S_{ki}$ is contained in N_{ji}, as is implied by the first formula in this group, then (5.5.2) must be true.

The first formula, on which the others are thus seen to depend, follows from $S_{jk}S_{ki} \subset S_{ji}$ if we remember that according to theorem (4.7.1) the product $(N_{ij}S_{jk})S_{ki}$ is associative.

We define now the set N which we claim to be the radical of the ring.

[5.6] The set of all elements whose components are singular with respect to the Peirce decomposition under discussion is denoted by N; in other terms $N = (i, k = 0 \cdots n) \sum N_{ik}$.

As announced in the beginning of this section, we have the theorem:

(5.7) *The set N is an ideal.*

First of all we see that by virtue of (5.4) the set N is a modulus; it remains to show that the products NS and SN are contained in N. In accordance with the rules (4.5) we have

$$NS = (\sum N_{ik})(\sum S_{lm}) = (i, j, k = 0 \cdots n) \sum N_{ij}S_{jk} + (i \neq k) \sum N_{ik}S_{ik};$$

the relations (5.5) have been derived in order to make evident that the last member is contained in N; the other relation $SN \subset N$ is verified in the same fashion.

We note that the ideal N does not contain any one of the elements e_i (these are regular) and is therefore different from the ring S. This would allow us to state another special instance of the final theorem, namely:

(5.8) *If a hypercomplex ring is simple and does not consist exclusively of nilpotent elements, then the only properly nilpotent element is zero. A fortiori, the 0-ideal is the radical of the ring.*

This partial result is of value in the theory of finite alternative division algebras.[22]

6. By means of the following lemma we show that every properly nilpotent element is contained in the ideal N.

(6.1) *If one component, say x_{ki}, of an element x is not singular, then we can find a left multiple $ax = a_{ik}x$ which is not nilpotent.*

Suppose that there is an element a_{ik} such that $a_{ik}x_{ki}$ is equal to e_i; the existence of such a quantity will be shown presently. We say that $a_{ik}x$ cannot be nilpotent, since the formula (4.8.3) makes it evident that the component in S_{ii} of this product and any of its powers is equal to e_i. In order to prove that the element a_{ik} exists we note first that none of the indices is zero, since x_{ki} was regular. If the indices are equal, then x_{ii} has an inverse in the ring S_{ii} which solves the equation. If finally i and k are different, we choose first an element y_{ik} such that $y_{ik}x_{ki}$ is not nilpotent; the latter product has again an inverse z_{ii} in the ring S_{ii}, so that $z_{ii}(y_{ik}x_{ki}) = e_i$. Since the last product is associative, by virtue of (4.7.1), we see that $a_{ik} = z_{ii}y_{ik}$ is the element we wanted.

[22] See (5) p. 402.

In the other direction we prove:

(6.2) *Every element of a modulus N_{ik} is properly nilpotent.*

Indeed, according to the formulas (4.8.1) and (4.8.2) the power $(a_{ik}x)^m$ is a sum of at most two terms which contain the factor $(a_{ik}x_{ki})^{m-1}$; the latter quantity vanishes for sufficiently high exponents, and we see that every multiple of a_{ik} is nilpotent. This implies in particular that the elements of the ideal N are at least sums of nilpotent elements, and we state less if we say:

(6.2.1) *The ideal N is contained in every ideal which contains all properly nilpotent elements.*

Combining this with theorem (6.1), which guarantees that every properly nilpotent element is necessarily a member of N, we obtain:

(6.2.2) *The ideal N is exactly the ideal which is generated by the properly nilpotent elements of the ring.*

This remark makes it clear that the ideal does not depend on the particular Peirce decomposition which has been used in the construction. We have mentioned before that the idempotents e_i are not elements of the ideal, since their components in S_{ii} are not singular; the same holds naturally for any sum of the form $(i = 1 \cdots m) \sum e_i$, provided that it is not zero. By virtue of (4.4) we may choose a Peirce decomposition for which a preassigned idempotent e turns out to be a sum of this type, and that implies:

(6.3) *The ideal N has no idempotent elements.*

If an ideal does not contain any idempotents then all its elements are nilpotent; for an element which is not nilpotent has an idempotent multiple; and an ideal contains the multiples of its elements. It is obvious, for the same reason, that all multiples of any element in the ideal are nilpotent. Thus we have finally arrived at the conclusion:

(6.4) *The ideal N, like any ideal of a hypercomplex ring without idempotents, consists entirely of properly nilpotent elements.*

If we combine the theorems (3.13), (6.1), (6.4) and the definitions [3.4] and [3.10] we may finally state the

(6.5) THEOREM. *Every hypercomplex alternative ring has a radical.*

Conclusion. If one is willing to restrict the class of hypercomplex rings the theory becomes much simpler. Mr. Artin has proved the existence of the radical for finite algebras of characteristic zero by extending the theory of traces.[23] By the use of an entirely different method (developed[24] first for Lie rings) I have been able to establish the nilpotency of the radical for alternative systems which satisfy the ascending chain condition for subsystems; this takes care of arbitrary finite algebras and likewise of rings which have a finite number of elements. Whether this fact is also a consequence of our three chain conditions I do not know.

[23] Not published; see (5) p. 402.

[24] See (6).

On the other hand it seems very probable that the three chain conditions are not the weakest assumptions which furnish a radical. Several intermediate results can be established under conditions which are compatible with infinite Peirce decompositions; and it seems that there is a less restrictive theory where infinite matrices with a finite number of nonvanishing rows play a role.

UNIVERSITY OF CALIFORNIA AT
LOS ANGELES

BIBLIOGRAPHY

1. M. Zorn, *Theorie der alternativen Ringe*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 8 (1930), p. 123.
2. E. Artin, *Zur Theorie der hyperkomplexen Zahlen*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 5 (1927), p. 251.
3. L. E. Dickson, *Algebras and their arithmetics*, Chicago 1923.
4. R. Moufang, *Zur Struktur von Alternativkörpern*, Mathematische Annalen, vol. 110 (1934), p. 416.
5. M. Zorn, *Alternativkörper und quadratische Systeme*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 9 (1933), p. 395.
6. M. Zorn, *On a theorem of Engel*, Bulletin of the American Mathematical Society, vol. 43 (1937), p. 401.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

SEPARATION SPACES[0]

BY A. D. WALLACE

(Received July 17, 1940)

1. Introduction

The purpose of this paper is to give a system of axioms suitable for the construction of a theory of connectivity in the large and for the study of connectivity invariants. The postulate system, which is of an essentially new type, is based on the notion of separation as the undefined concept. In sections 2 and 3 we correlate our system with the usual topological concepts, in section 4 proofs are given for the "expected" theorems, and sections 5 and 6 are devoted to some general examples. Section 7 is devoted to transformations and a generalization of the Eilenberg-Whyburn theorem on the factorization of continuous transformations.

2. Axioms and Definitions

We envisage an abstract set of elements S together with a binary relation $X \mid Y$ between subsets X and Y of S. By a *weak separation-space* (or an s_w-space) we mean a system $(S, |)$ in which the following axioms hold:

AXIOM S.I. *The null set is separated from every set*: $(X \subset S) \rightarrow (0 \mid X)$.[1]

AXIOM S.II. *Separation is symmetric*: $(X \mid Y) \rightarrow (Y \mid X)$.

AXIOM S.III. *Separation is disjunctive*: $(X \mid Y) \rightarrow (X \cdot Y = 0)$.

AXIOM S.IV. *Separation is hereditary*: $(X \mid Y) \cdot (X_1 \subset X) \rightarrow (X_1 \mid Y)$.

For convenience we adopt the conventions:

(i) The symbol "$X \mid Y$" is to be read "*X is separated from Y*" or "*X and Y are separated.*"

(ii) In a topological space two sets X and Y will be said to be "*topologically separated*" if they satisfy the condition of Lennes-Hausdorff, $\bar{X} \cdot Y + X \cdot \bar{Y} = 0$.

Before examining particular instances of s_w-spaces it will be necessary to have the

(2.1) DEFINITION. *If X is a subset of an s_w-space then kX is the set of all points no one of which is separated from X: $(p \in kX) \equiv (p \mid X)'$. By $H(X)$ is meant the class of all sets Y such that $Y = kY \supset X$, and hX is the set of points common to all the sets in the class $H(X)$.*

The topology arising from the adoption of kX (respectively hX) as the *closure* of the set X will be called the *k-topology* (respectively *h-topology*) of the s_w-space.

[0] Presented to the Am. Math. Soc., April 8, 1939 and April 26, 1940.

[1] We adopt the logical notation of Kuratowski [1]. Thus if A and B are propositions then $A \equiv B$, $A \rightarrow B$, $A + B$, $A \cdot B$, A' are respectively A is equivalent to B, A implies B, A or B, A and B, and A is false. The symbol $(X \mid Y)'$ will be constantly used.

(2.2) Example. Let S be any set of elements and let "$X \mid Y$" mean that X and Y are disjoint. Here the system $(S, \mid)$ is an s_w-space in which $X = kX = hX$ so that the k-topology and the h-topology are each discrete.

(2.3) Example. Let S be a topological space in which *closure* satisfies the axioms *K.I–K.III* of (3.3), i.e., S is a T_1-space. Define "$X \mid Y$" to mean that X and Y are topologically separated, $\bar{X} \cdot Y + X \cdot \bar{Y} = 0$. It is clear that the system $(S, \mid)$ is a weak separation-space with $\bar{X} = kX = hX$.

(2.4) Example. In this example S is Euclidean n-space and "$X \mid Y$" is to mean the existence of an $(n - 1)$-dimensional subspace L such that X and Y lie in different open half-spaces determined by L. Here we have $\bar{X} \subset hX = kX$, and this latter set is the closed convex hull of X. In this connection see §4.

(2.5) Theorem. *In an s_w-space we have*

K.I_w If $X \subset Y$ then $kX \subset kY$.

K.II_w $X \subset kX, \quad k0 = 0$.

H.I_w If $X \subset Y$ then $hX \subset hY$.

H.II_w $X \subset hX, \quad h0 = 0$.

H.III_w $hhX = hX$.

HK $X \subset kX \subset hX = hkX = khX$.

Proof: (K.I_w). If $x \in kX$ then $(x \mid X)'$ and hence by S.IV it follows that $(x \mid Y)'$ so that $x \in kY$.

(K.II_w). For the first part, if $x \in X$ then $(x \mid X)'$ by S.III so that $x \in kX$. If $x \in k0$ then $(x \mid 0)'$ and hence $(0 \mid x)'$ by S.II and this is a contradiction to S.I.

(H.I_w). Using the notation of (2.1) we have to show that if $X \subset Y$ then $H(Y) \subset H(X)$. If $R \in H(Y)$ then $X \subset Y \subset R = kR$. From this it follows that $R \in H(X)$.

(H.II_w). For the first part it suffices to show that $kX \subset hX$ by virtue of K.II_w. But if $x \in kX$ it follows that $(x \mid R)'$ for any $R \in H(X)$, so that $x \in kR = R$. Thus x is contained in every element of $H(X)$ and thus $x \in hX$. For the second part, since $k0 = 0$ we have $0 \in H(0)$ and hence $h0 = 0$.

Before continuing with the proof of (2.5) we need the

(2.6) Definition. *If X is a subset of an s_w-space and $X = kX$ then X is said to be k-closed.*

(2.7) Theorem. *The common part of any aggregate of k-closed sets is k-closed.*

Proof: Let X be the product of the family of sets $[X_\alpha]$ where $X_\alpha = kX_\alpha$. By K.I_w we have $kX \subset kX_\alpha = X_\alpha$ so that $kX \subset X$. In consequence of K.II_w we have $X = kX$.

Returning to the proof of (2.5): (H.III_w). We need only show that $hhX \subset hX$. By definition, hX is k-closed, i.e., we have $khX = hX$. But $hX \subset hX = khX$ and this qualifies hX for admission to the class $H(hX)$, and so $hhX \subset hX$.

(HK). We have proved thus far that

$$X \subset kX \subset hX = khX.$$

From the first inclusion and H.I_w we get $hX \subset hkX$ and we know that $hkX \subset hhX = hX$ in consequence of K.II_w and H.I_w. From this we have $hX = hkX$ and thus the proof of (2.5) is complete.

If the operator h (or k) is to be an analog of closure it should be additive. This requires that separation be additive. An s_w-space satisfying

AXIOM S.V. *Separation is additive*: $(X_1 \mid Y)\cdot(X_2 \mid Y) \rightarrow ((X_1 + X_2) \mid Y)$, will be called a *separation space* or an *s-space*.

Examples (2.2) and (2.3) are s-spaces while Example (2.4) is not. For if S is the Euclidean plane and X_1, Y and X_2 are respectively the points with coordinates $(-1, 0)$, $(0, 0)$ and $(1, 0)$ it is clear that S.V does not hold.

(2.8) THEOREM. *In an s-space* K.I$_w$ *and* H.I$_w$ *can be replaced by the stronger properties*

K.I $k(X + Y) = kX + kY$

H.I $h(X + Y) = hX + hY$.

PROOF: It follows readily from K.I$_w$ and H.I$_w$ that we have $kX + kY \subset k(X + Y)$ and $hX + hY \subset h(X + Y)$. If $x \in k(X + Y)$ this means that $(x \mid (X + Y))'$ and hence by S.II it follows that $((X + Y) \mid x)'$ and so by S.V either $(X \mid x)'$ or $(Y \mid x)'$. From this, using S.I, it is manifest that $x \in kX + kY$, proving K.I. To prove H.I we assume that $x \in h(X + Y)$ but is not a point of $hX + hY$. Then x is contained in every R in the class $H(X + Y)$ but there exist sets R_1 and R_2, respectively elements of $H(X)$ and $H(Y)$, which do not contain x. Now by K.I we have

$$R_1 + R_2 = kR_1 + kR_2 = k(R_1 + R_2) \supset X + Y$$

and x is not a point of $R_1 + R_2$. This a contradiction since $(R_1 + R_2) \in H(X + Y)$.

3. Topological Characterization of S-Spaces

We begin by stating

AXIOM S.VI. *If* $x \mid X$ *then* $x \mid kX$.

AXIOM S.VII. *If* x *and* y *are distinct points then* $x \mid y$.

An s-space satisfying the above two axioms will be called an *s_k-space*. Examples (2.2) and (2.3) are s_k-spaces. While axioms S.VI and S.VII hold in Example (2.4) it is not an s_k-space since it is not an s-space. The effect of these new postulates is shown by the

(3.1) THEOREM. *In an s_k-space the conditions* K.I–K.III *of* (3.3) *hold and* $kX = hX$.

PROOF: It is clear that S.VII and (2.5) imply K.II. In consequence of K.II$_w$ and K.I of (2.5) and (2.8) we need only show that $kkX \subset kX$. But this is immediate from S.VI. Further K.III implies $kX \in H(X)$ and so $hX \subset kX$. From HK of (2.5) we get equality here.

The fact that K.I–K.III of (3.3) hold in an s_k-space does not imply that $X \mid Y$ is equivalent to $X\cdot kY + Y\cdot kX = 0$.

(3.2) EXAMPLE. Let S be the unit interval with its usual topology and define "$(X \mid Y) \equiv (\bar{X}\cdot\bar{Y} = 0)$". The system $(S, |)$ is an s_k-space with $\bar{X} = hX = kX$. If P and Q denote respectively the sets $0 \leqq x < 1/2$ and $1/2 < x \leqq 1$ then $(P \mid Q)'$ but they satisfy the condition of Lennes-Hausdorff since they are top-

ologically separated. It is interesting to remark that S has no s-cut-points and indeed any subset R of S which is dense in S is s-connected (cf. §4).

This example indicates the need for the following

Axiom S.VIII. *If for each $x \epsilon X$ and each $y \epsilon Y$ we have $(x \mid Y)$ and $(y \mid X)$ then $(X \mid Y)$.*

An s_k-space satisfying S.VIII will be called an *s_t-space.*

(3.3) Characterization Theorem. *In an s_t-space we have*

K.I $k(X + Y) = kX + kY$.

K.II $kX = X$ *if X is null or a point.*

K.III $kkX = kX$.

R.I $(x \epsilon kX) \equiv (x \mid X)'$.

R.II $(X \mid Y) \equiv (X \cdot kY + Y \cdot kX = 0)$.

Conversely, if the operator k satisfies K.I–K.III *and separation is defined by* R.II *then* S.I–S.VIII *and* R.I *hold.*

Proof: It remains only to show, in consequence of (3.1) and the definitions, that R.II is valid. Suppose that the left member of R.II holds. If $X \cdot kY \neq 0$ (say) then there would be a point in X not separated from Y and hence $(X \mid Y)'$. If the right member holds and the left does not, then by S.VIII we can find a point $x \epsilon X$ such that $(x \mid Y)'$. That is, $X \cdot kY \neq 0$, a contradiction. It is readily seen that the converse part of the theorem is valid.

It is to be noted that there exist spaces in which closure satisfies K.I, K.II$_w$ of (2.5), and K.III and in which separation (as defined by R.II) does not lead to R.I. This is simply because in such a space a point need not be a closed set, i.e., $kx \neq x$. Nevertheless we have

(3.4) Theorem. *In an s-space, for any point x and any set X*

$$(x \mid X) \equiv (X \cdot kx + x \cdot kX = 0).$$

Proof: The right member manifestly implies the left. Assume that the right member is false and first suppose that $x \cdot kX \neq 0$. Then by definition $(x \mid X)'$ and hence the left member fails to hold. Suppose only that $X \cdot kx \neq 0$. Then there is a $p \epsilon X$ which is not separated from x. Consequently we have $(x \mid p)'$ and hence $x \epsilon kp \subset kX$ in virtue of K.I$_w$ of (2.5). Again the left member fails to hold.

Topological spaces of the type (3.3) were studied extensively by Kuratowski [1] and others and are called T_1-spaces by Alexandroff-Hopf [1]. They may be described in terms of closed sets by saying that (i) the null set, the space and each point is closed (ii) the sum of two and the common part of any number of closed sets is again a closed set.

4. S-Connected Sets

In this section we assume that S *is an s-space* (Axioms S.I–S.V). Our purpose is to demonstrate that the "expected" theorems hold concerning "connected" sets.

(4.1) Definition. *A subset of an s-space is said to be s-connected provided that it is not the sum of two non-null separated sets.*
(4.2) Theorem. *If X is s-connected and contained in $Y + Z$, $Y \mid Z$, then $X \subset Y$ or $X \subset Z$.*

Proof: If X intersects both Y and Z then $X = X \cdot Y + X \cdot Z$ and neither summand is null. By S.II, S.IV, and S.V it follows that $X \cdot Y \mid X \cdot Z$ and this is a contradiction to the definition of s-connectivity.
(4.3) Theorem. *If $[X_\alpha]$ is a class of s-connected sets no one of which is separated from the s-connected set X, then the set $X_0 = X + \sum X_\alpha$ is s-connected.*

Proof: If not, then $X_0 = Y + Z$ and $Y \mid Z$, neither being null. Then using (4.2) we may suppose that $X \subset Y$. Since $Z \cdot X_\alpha \neq 0$ for some α we have (again by (4.2)) that some X_α is contained in Z. By S.II, S.IV and S.V this implies that $X \mid X_\alpha$, a contradiction.
(4.4) Theorem. *If X is a non-null s-connected subset of the set Y then there is a maximal s-connected subset X_0 of Y which contains X. The set X_0 is called the s-component of Y which contains X.*

Proof: It is readily verified that the set $X_0 = X +$ [*all s-connected subsets of Y which are not separated from X*] is an s-connected subset of Y containing X and is not a proper subset of any other s-connected subset of Y having these properties. Hence X_0 is the desired set.
(4.5) Theorem. *If Y is an s-connected subset of the s-connected set X then*
(i) *If $X - Y = P + Q$, $P \mid Q$, then $Y + P$ and $Y + Q$ are s-connected;*
(ii) *If Z is an s-component of $X - Y$ then $X - Z$ is s-connected.*

Proof: (i). If we assume that $Y + P = M + N$ where $M \mid N$ and neither summand is null it follows by (4.2) that we may assume that Y is contained in N and hence by S.III that M is a subset of P. Hence, since,

$$X = (X - Y) + Y = P + Q + Y = M + (N + Q),$$

on applying S.IV and S.V it follows that $M \mid (N + Q)$, a contradiction to the assumption that X is s-connected.
(ii) Assume that $X - Z = M + N$ where $M \mid N$. Since Y does not intersect Z and Y is s-connected we may assume that Y is a subset of N. By (i) the set $M + Z$ is an s-connected subset of $X - Y$ and since Z is an s-component of this set and $M \cdot Z = 0$ we have $M = 0$.

A somewhat different form of the first part of (4.5) is as follows:
(4.6) Theorem. *If P and $P \cdot Q$ are s-connected and $R \mid (P - (Q + R))$ then $P \cdot (Q + R)$ is s-connected.*
(4.7) Theorem. *If $X + Y$ and $X \cdot Y$ are s-connected and $(X - Y) \mid (Y - X)$ then X and Y are s-connected.*

Proof: We have

$$(X + Y) - X \cdot Y = (X - Y) + (Y - X) \quad \text{and} \quad X = X \cdot Y + (X - Y)$$

so that we may apply (4.5).

In (4.7) the condition $(X - Y) \mid (Y - X)$ replaces the classical (and un-

necessarily restrictive) "X and Y are closed in $X + Y$" which indeed is not applicable in an s-space. To show the efficacy of the condition it may be shown, for example, that *in order that a connected, locally connected metric space R be unicoherent it is necessary and sufficient that for any decomposition $R = P + Q$ where P and Q are connected and $(P - Q) \mid (Q - P)$ (topologically) we have $P \cdot Q$ connected.*

(4.8) THEOREM. *If X is s-connected and $X \subset X_0 \subset kX$ then X_0 is s-connected.*

PROOF. The set X_0 is of the form $X + \sum X_\alpha$ where each X_α is a point of $kX - X$. We then apply (4.3).

(4.9) THEOREM. *If X is s-connected so is hX.*

PROOF: From (4.8) we know that kX is s-connected. If $hX = M + N$ where $M \mid N$ then, because X is a subset of hX, we may assume that X is contained in M. Since $M \mid N$ it follows that $M \cdot kN = 0 = N \cdot kM$. But

$$M + N = hX = khX = k(M + N) = kM + kN$$

and thus $M = kM \supset X$ so that $hX \subset M$ and therefore $hX = M$. Hence $N = 0$ and we see that hX is s-connected.

If the set I is s-connected and contains the points a and b and no proper subset of I enjoys both of these properties then I is said to be *irreducibly s-connected between a and b*. It may be shown that if $x \in I$, $a \neq x \neq b$ then $I - x = I_1 + I_2$ where $a \in I_1$, $b \in I_2$, $I_1 \mid I_2$ and both I_1 and I_2 are connected. This enables one to order the elements of I (from a to b) by defining $p < q$ if there is an $x \in I$ giving the decomposition above with $p \in I_1$ and $q \in I_2$. We state without proof the[2]

THEOREM. *If I is an s-space satisfying Axiom* S.VIII *and I is irreducibly s-connected between a and b, then, if the elements of I are ordered as above, the Dedekind proposition holds.*

5. An Example

Let S be a space satisfying the first three axioms of Kuratowski (i.e., *closure* satisfies K.I–K.III of (3.3)) and let $G = [Z]$ be a family of closed sets of S. Throughout this section we *define* separation as follows: *The sets X and Y are separated* (i.e., $X \mid Y$) *provided that one of them is null or there exists an element Z of G such that*

$$(+) \quad S - Z = U + V, \qquad X \subset U, \qquad Y \subset V, \qquad \bar{U} \cdot V + U \cdot \bar{V} = 0.$$

(5.1) THEOREM. *The system $(S, \mid)$ is an s_w-space and for any subset P of S the sets kP and hP are closed and $\bar{P} \subset kP \subset hP$.*

PROOF: It is manifest that the $(S, \mid)$ is an s_w-space. We proceed to show that kP is closed. To this end it suffices to prove kP is the common part of a collection of sets $H = [U + Z]$ satisfying $(+)$. We define this collection as follows: For any x not in kP we know that $(+)$ holds with $P = X$ and $x = Y$ and we

[2] See Knaster and Kuratowski [1].

let H be the collection of all sets $U + Z$ thus obtained. If we let Q be the product of all the sets in H then $kP \subset Q$. The reverse inclusion is easily seen to be valid.

If the collection G contains only the null set then for each X in S the set kX is the sum of all the quasi-components of S which intersect $\bar{X}$. If G contains all closed sets and if further S has the property that for any two topologically separated subsets of S the condition (+) holds, then the system $(S, |)$ is an s_t-space and $X = kX = hX$. Further, condition R.II of (3.3) is valid so that separation and topological separation are the same. Thus varying the collection G enables us to get anything from an s_w-space to the full topology of S.

For the remainder of this section we suppose that the family G *is a multiplicative ideal in the lattice of closed subsets of* S, that is, if Z_1 and Z_2 are in G then so is $Z_1 + Z_2$, and if Z_0 is a closed subset of an element of G then $Z_0 \in G$.

(5.2) Lemma. *If for* $i = 1, 2, \cdots n$ *we have*

$$(+)_i \qquad S - Z_i = U_i + V_i, \quad X_i \subset U_i, \quad Y_i \subset V_i,$$

$$\bar{U}_i \cdot V_i + U_i \cdot \bar{V}_i = 0, \quad Z_i \in G,$$

then there is a $Z \in G$ *such that* (+) *holds with* $X = X_1 \cdot X_2 \cdot \cdots \cdot X_n$ *and* $Y = Y_1 + Y_2 + \cdots + Y_n$.

Proof: Since the general case follows by induction we restrict our attention to the case $n = 2$. Let $U = U_1 \cdot U_2$ and $V = V_1 + V_2$, and set $Z = S - (U + V)$. That $\bar{U} \cdot V + U \cdot \bar{V} = 0$ is immediate and clearly $X \subset U$ and $Y \subset V$. Further we have

$$\begin{aligned} Z = (S - U) \cdot (S - V) &= [(S - U_1) + (S - U_2)] \cdot [S - V] \\ &= [(Z_1 + V_1) + (Z_2 + V_2)] \cdot [S - V_1] \cdot [S - V_2] \\ &= (Z_1 - V_2) + (Z_2 - V_1). \end{aligned}$$

Since V_1 and V_2 are open it follows (since G is an ideal) that $Z \in G$.

(5.3) Theorem. *The system* $(S, |)$ *is an s-space.*

Proof: In virtue of (3.1) we need only prove that S.V holds. But if $X \mid Y_1$ and $X \mid Y_2$ then $(+)_i$ holds for $i = 1, 2$ and $X = X_1 = X_2$. Consequently (+) holds with $Y = Y_1 + Y_2$.

The condition that G be an ideal is essential here, and even in this event it is possible to give examples of Peano spaces in which S.VI–S.VIII all fail to hold

(5.4) Lemma. *If* P *is closed,* S *is compact metric, and for each* $p \in P$ *we have* $p \mid Q$, *then* $P \mid Q$.

Proof. For each $p \in P$ condition (+) holds with $X = Q$ and $p = Y$. Since B is compact and each V is open we can find an integer n such that $(+)_i$ holds for $i = 1, 2, \cdots n$, and $P \subset V_1 + V_2 + \cdots + V_n$. Thus (+) holds with $X = Q$ and $Y = P$ and this completes the proof.

(5.5) THEOREM. *If S is compact metric and P is s-connected then kP and hP are connected.*

PROOF: If kP is not connected then we have $kP = M + N$ where M and N are disjoint closed sets, neither of which is void. There exists a closed subset F such that

$$S - F = A + B, \qquad M \subset A, \qquad N \subset B, \qquad \bar{A}\cdot B + A\cdot\bar{B} = 0.$$

Since $F\cdot kP = 0$ we have $f \mid P$ for each $f \in F$ and thus by (5.4) we conclude that $F \mid P$. Thus (+) is valid with $X = P$ and $Y = F$. Since $P \subset A + B$ we may suppose that $P\cdot A \neq 0$, and thus $P\cdot(U - \bar{B}) \neq 0$ in virtue of the fact that $A\cdot\bar{B} = 0$. Further since $\bar{B} - B \subset F \subset V$ we have $U - B = U - \bar{B}$. It is clear that the sets $U - B$ and $V + B$ are topologically separated being disjoint and open, and moreover $S - Z = (U - B) + (V + B)$. But since P is s-connected and $P\cdot(U - B)$ is not null we have $P \subset U - B$. Further $N\cdot(V + B)$ is not null and so for some $q \in N \subset kP$ we have $q \mid P$, a contradiction. The proof that hP is connected is as follows: By (4.9) hP is s-connected and by (2.5) $khP = hP$ so that by what we have just shown the set hP is connected.

The following example has many interesting features.

(5.6) EXAMPLE. Let S be the unit interval with its usual topology, let s be the point 0, and let G consist of all closed subsets of the sequence $(1, 1/2, 1/3, \cdots, 1/n, \cdots 0)$. Notice that $s = hs = ks$ but that if X is a closed subset which does not contain the point s then $kX \neq s$ while $LX = s$. Here the only k-closed (h-closed) sets are the null set, the point s and S.

While the systems studied in this section have long been known in topology, going back to Poincaré (cf. Menger [1, 2], Moore [1]), I am not aware that any effort has been made to introduce a topology of separation, or to superimpose a topology of any sort on the system. The following proposition has a certain intrinsic interest.

(5.7) THEOREM. *The relationship of "not being separated" is a continuous function on the space of all closed sets of S in the sense that if $X_n \rightarrow X$ and $Y_n \rightarrow Y$, then $(X_n \mid Y_n)'$ implies $(X \mid Y)'$, S being assumed compact and metric.*

6. A Dynamical Example

In the present section we borrow the terminology of dynamics (cf. Hedlund [1]). Roughly speaking two sets will be "separated" if they are not "joined" by a motion. More exactly we consider a metric space S and a family $T = [t]$ of subsets of S (not necessarily closed). Further we suppose that through each point of S there passes at least one *motion* (= element of T). A subset P of S will be *invariant* provided that $P = \sum t,\ t\cdot P \neq 0$.

(6.1) DEFINITION. *The subsets X and Y of S will be separated* (i.e., $X \mid Y$) *provided that*

(i) *No element of T intersects both $\bar{X}$ and Y.*

(ii) *No element of T intersects both X and $\bar{Y}$.*

If the elements of T are points of S then separation and topological separation

are equivalent and the k-topology and topology of S agree. It may be seen that the following is valid.

(6.2) THEOREM. *The system* $(S, |)$ *is an s-space in which Axiom* S.VIII *holds. For any subset* X *of* S *we have*

$$kX = \sum t, t \cdot \bar{X} \neq 0.$$

In order that a set be k-closed it is necessary and sufficient that it be closed and invariant. The set hX is the smallest closed and invariant set containing X.

In general a set hx ($x \in S$) will not be a minimal set (in the sense that it may contain a proper k-closed subset). It is readily seen that if the motions are connected sets then if P is s-connected sets kP and hP are also s-connected and connected.

7. Transformations and the Eilenberg-Whyburn Theorem

If A and B are s-spaces then the single-valued transformation $T(A) = B$ will be called *s-continuous* provided that

$$(X \mid Y) \equiv (T^{-1}(X) \mid T^{-1}(Y))$$

for each pair of subsets X, Y of B.

This definition is justified by the

(7.1) THEOREM. *If A and B are compact metric spaces and "separation" is interpreted to mean "topological separation" then* $T(A) = B$ *is continuous if and only if it is s-continuous.*

PROOF: If T is continuous it is readily seen that it is s-continuous. If T is not continuous we can find a sequence $x_n \to x$ in A such that $T(x_n)$ converges to a point $y \neq T(x)$. It follows that we may suppose that the sets $T(x)$ and $T(x_1) + T(x_2) + \cdots$ are separated. Hence the sets x and $x_1 + x_2 + \cdots$ will be separated in contradiction to the assumption that x is an accumulation point of the set $x_1 + x_2 + \cdots$.

The following proposition (while not immediately relevant to our purpose) has considerable intrinsic interest.

(7.2) THEOREM. *If* $T(A) = B$ *is continuous where A and B are compact metric then the following properties are equivalent ("separation" meaning "topological separation"):*

(i) *The image of an open set is open.*

(ii) *If* $X \subset A$ *and* $Y \subset B$ *then* $(X \mid T^{-1}(Y)) \equiv (T(X) \mid Y)$.

PROOF: Note first that (under our convention above) a set V is open if and only if $p \mid (B - V)$ for each $p \in V$. Hence if T satisfies (ii) it is to be shown that if U is open then $y \in T(U)$ implies $y \mid (B - T(U))$. Now if $x \in T^{-1}(y) \cdot U$ then x and $A - U$ are separated and hence the sets x and $A - T^{-1}T(U) = T^{-1}(B - T(U))$ are separated (in virtue of S.IV). It then follows by (ii) that $y \mid (B - T(U))$. We suppose now that T satisfies (i). To prove the leftward implication of (ii) we see that

$$(X \mid T^{-1}(Y))' \to (T^{-1}T(X) \mid T^{-1}(Y))' \to (T(X) \mid Y)'$$

by S.IV and (7.1). To prove the rightward implication we note that if (ii) holds then $T^{-1}(\bar{P}) = \overline{T^{-1}(P)}$ for each $P \subset B$, and $T(\bar{Q}) = \overline{T(Q)}$ for each $Q \subset A$. Hence

$$(X \mid T^{-1}(Y)) \rightarrow (\bar{X}\cdot T^{-1}(Y) + X\cdot T^{-1}(\bar{Y}) = 0)$$

$$\rightarrow (\overline{T(X)}\cdot Y + T(X)\cdot \bar{Y} = 0) \rightarrow (T(X) \mid Y).$$

We suppose from now on that A is an s-space and $T(A) = B$ is single-valued and varying hypotheses will be placed on T and B.

(7.3) *If B is any abstract set of elements and we define for each pair X, Y in B*

$$(X \mid Y) = (T^{-1}(X) \mid T^{-1}(Y)$$

then the system $(B, |)$ is an s-space and T is s-continuous

PROOF: To validate the second statement it remains only to prove the first. Since S.I–S.III are immediate we have to show that S.IV and S.V hold. But these are immediate in virtue of the identities $(P \subset Q) \equiv (T^{-1}(P) \subset T^{-1}(Q))$ and $T^{-1}(P + Q) = T^{-1}(P) + T^{-1}(Q)$, and the axioms.

(7.4) THEOREM. *If A' and B are s-spaces, $T_1(A) = A'$ and $T_2(A') = B$ are singled-valued, and $T = T_2T_1$ and T_1 are s-continuous, then T_2 is s-continuous.*

PROOF: If X and Y are subsets of B and $X \mid Y$, then we have $T^{-1}(X) \mid T^{-1}(Y)$. But $T^{-1}(X) = T_1^{-1}T_2^{-1}(X)$ and $T^{-1}(Y) = T_1^{-1}T_2^{-1}(Y)$ so that since T_1 is s-continuous we have $T_2^{-1}(X) \mid T_2^{-1}(Y)$. The reverse implication follows similarly so that T_2 is s-continuous.

(7.5) THEOREM. *If $T(A) = B$ is s-continuous (B an s-space) and $T^{-1}(b)$ is s-connected for each $b \in B$, then $T^{-1}(B_0)$ is s-connected for each s-connected subset B_0 of B.*

PROOF: Suppose that $T^{-1}(B_0) = P + Q$ where $P \mid Q$ and neither summand is void. Then for any $b \in B_0$ we have $T^{-1}(b)$ contained in either P or Q. Thus $P = T^{-1}T(P)$ and $Q = T^{-1}T(Q)$ so that we get $T(P) \mid T(Q)$. Since $B_0 = T(P) + T(Q)$ it follows that B_0 is not s-connected, a contradiction.

The following is a generalization of a theorem due to Eilenberg [1] and Whyburn [1]. It is to be noted that no use is made of upper semi-continuous decompositions in the proof of the theorem.

(7.6) THEOREM. *If A and B are s-spaces and $T(A) = B$ is s-continuous then there exists an s-space A' and s-continuous transformation $T_1(A) = A'$ and $T_2(A') = B$ such that $T = T_2T_1$ where*

(i) *for each $a' \in A'$ the set $T_1^{-1}(a')$ is connected and*

(ii) *for each $b \in B$ the set $T_2^{-1}(b)$ contains no s-connected set containing more than one point.*

PROOF: Let $A' = [a']$ be the class consisting of all s-components of all sets $T^{-1}(b)$, $b \in B$. For each $a \in A$ let $T_1(a) = a'$ be the s-component of $T^{-1}T(a)$ which contains a. For $X', Y' \subset A'$ define $X' \mid Y'$ to mean that $T_1^{-1}(X') \mid T_1^{-1}(Y')$. Then by (7.3) T_1 is s-continuous and A' is an s-space. By (7.4) $T_2 = T_1T^{-1}$ is s-continuous. If X' is an s-connected subset of $T_2^{-1}(b)$, $b \in B$,

then by (7.5) we know that $T_1^{-1}(X')$ is s-connected and is contained in $T_1^{-1}T_2^{-1}(b) = T^{-1}(b)$. Thus $T_1^{-1}(X')$ is in a single component of $T^{-1}(b)$ and thus X' is a single point of A'.

THE UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY

Alexandroff-Hopf........................[1], *Topologie* I, Berlin.
Eilenberg...............................[1], Fund. Math. 22(1934), p. 292.
Hedlund................................[1], Bull. Am. Math. Soc. 45(1939), p. 241.
Knaster and Kuratowski................[1], Fund. Math. 2(1921), p. 206.
Kuratowski.............................[1], *Topologie* I, Warsaw.
Menger.................................[1], *Dimensionstheorie*, Berlin.
Menger.................................[2], *Kurventheorie*, Berlin.
Moore...................................[1], *Foundations of Point Set Theory*, New York.
Whyburn...............................[1], Am. Journ. Math. 56(1934), p. 370.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

TENSOR DECOMPOSITION WITH APPLICATIONS TO THE CONTACT AND COMPLEX GROUPS

BY T. C. DOYLE

(Received September 12, 1939)

INTRODUCTION

The original purpose of this paper was to apply tensor methods to a study of the invariant theory of homogeneous contact transformations. L. P. Eisenhart and M. S. Knebelman (1) have shown how a vector Λ of $2q$ components in the phase space of $2q$ coordinates x^i and p_i may be decomposed into two geometric objects λ and μ of q components each. The transformation laws of λ and μ were such that whereas the new components $\bar{\lambda}$ of one set were linear combinations of the corresponding old components λ, the new components $\bar{\mu}$ were linearly dependent on both sets λ and μ. Because of this transformation behavior the decomposition just cited will be described as *partial*. In this paper we shall arrive at a complete decomposition in the sense that the $\bar{\lambda}$'s will be defined in terms of the λ's only and the $\bar{\mu}$'s in terms of the μ's. The underlying formalism of the Eisenhart-Knebelman paper has been employed here and the decomposition at which we arrive agrees, when restricted to extended point transformations, with their similarly restricted decomposition.

Our decomposition has been accomplished by introducing the notion of two families of complementary subspaces of the phase space and of projection tensors in these subspaces. Since the method is independent of any implication of a contact transformation on the coordinates it will first be developed in its full generality and the decomposition under the contact group will be presented as a special application of the general theory. As additional applications we shall give two decompositions of tensors under the complex group, one being real and the other complex. In section 16 the doubly homogeneous contact group of Schouten (2) is derived independently and a complete decomposition for this group in terms of an affine connection is given.

1. THE UNDERLYING SYSTEMS OF DIFFERENTIAL EQUATIONS

Two systems of partial differential equations of the solved type

$$X_\alpha f = L^Q{}_\alpha \partial_Q f = \partial_\alpha f + L^r{}_\alpha(x^A)\,\partial_r f = 0 \tag{1.1}$$

$$X_i f = L^Q{}_i \partial_Q f = \partial_i f + L^\rho{}_i(x^A)\,\partial_\rho f = 0 \tag{1.2}$$

for which $|L^A{}_B| = \begin{vmatrix} \delta^\alpha{}_\beta & L^\alpha{}_j \\ L^i{}_\beta & \delta^i{}_j \end{vmatrix} \neq 0$ and $\partial_A f = \dfrac{\partial f}{\partial x^A}$ and wherein the indices have the ranges $(\alpha) = (1, \cdots, q)$, $(i) = (q+1, \cdots, N)$, and $(A) = (1, \cdots, N)$ will be called *complementary*.

Associated with the systems $X_i f = 0$ and $X_\alpha f = 0$ are their dual Pfaffian systems $P^\alpha = N^\alpha{}_Q\, dx^Q = dx^\alpha - L^\alpha{}_r\, dx^r = 0$ and $P^i = N^i{}_Q\, dx^Q = dx^i - L^i{}_\rho\, dx^\rho = 0$ with the matrix $\| N^A{}_B \| = \begin{Vmatrix} \delta^\alpha{}_\beta & -L^\alpha{}_j \\ -L^i{}_\beta & \delta^i{}_j \end{Vmatrix}$. From these definitions, if $f(x)$ is a solution of $X_i f = 0$ its differential $df = \partial_\alpha f(dx^\alpha - L^\alpha{}_r\, dx^r)$ will vanish for any increment dx^A satisfying $P^\alpha = 0$. Conversely, if $f(x)$ is an integral of $P^\alpha = 0$ then $df = (\partial_i f + L^\rho{}_i \partial_\rho f)\, dx^i$ must vanish identically in dx^i and $f(x)$ is a solution of $X_i f = 0$.

We shall confine our attention to (1.1) in making a few general remarks. Assuming for the moment complete integrability, there exist $N - q$ principal solutions (3) $\varphi^i(x^A)$ satisfying the initial conditions $\varphi^i(x_0^\alpha, x^j) = x^i$ within a suitably small neighborhood of a point $(x_0{}^A)$ for which $L^i{}_\alpha(x^A)$ are regular and the most general system of solutions is given by any $N - q$ independent functions $f^i(\varphi^j)$ of the principal solutions. Since $\partial_j \varphi^i \,|_{x^\alpha = x_0^\alpha} = \delta^i{}_j$ it follows that the determinant $| \partial_j f^i |$ does not vanish identically. A general system of solutions $f^i(x)$ determines an ∞^{N-q} family of q-dimensional subspaces $S^q(c^i)$ as defined by the equations $f^i(x) = c^i$. A parametric representation may be obtained by solving these equations to give $x^i = \psi^i(x^\alpha, c^j)$, the complete parametric system for $S^q(c^i)$ being

$$x^\alpha = x^\alpha \qquad x^i = \psi^i(x^\alpha, c^j). \tag{1.3}$$

Now the functions $F^i = x^i - \psi^i(x^\alpha, c^j)$ when equated to zero represent the same family of subspaces $S^q(c^i)$ as defined by $f^i(x) = c^i$ and hence must satisfy the system (1.1) for all values of x^A lying in $S^q(c^i)$, that is,

$$\partial_\alpha \psi^i = L^i{}_\alpha(x^\beta, \psi^j(x^\beta, c^k)). \tag{1.4}$$

Suppose $S^q(c^i)$ to be parameterized in a general manner by the equations $x^A = x^A(u^\alpha)$. Then the q invariants $\lambda_\alpha = \Lambda_Q \dfrac{\partial x^Q}{\partial u^\alpha}$ are the scalar products of Λ_A with the q tangent vectors $\dfrac{\partial x^A}{\partial u^\alpha}$ of $S^q(c^i)$. Under parameter transformations the λ's constitute a q-component vector of $S^q(c^i)$ which will be called the projection of Λ_A on $S^q(c^i)$. A distinguishing feature of the parameterization (1.3) may be emphasized by writing it in the form

$$x^\alpha = u^\alpha \qquad x^i = \psi^i(u^\alpha, c^j) \tag{1.5}$$

and transforming the x's by $\bar{x}^A = f^A(x)$ so that $\bar{x}^\alpha = f^\alpha(x^\beta, \psi^i(x^\beta, c^j))$. If we insist on invariance of form for the first set of (1.5), namely $\bar{x}^\alpha = \bar{u}^\alpha$, it may be seen that a transformation $x^A \to \bar{x}^A$ induces the parameter transformation $u \to \bar{u}$ given by $\bar{u}^\alpha = f^\alpha(u^\beta, \psi^i(u^\beta, c^j))$. The condition for non-singularity is $\left| \dfrac{\partial \bar{u}^\alpha}{\partial u^\beta} \right| = \left| \dfrac{\partial f^\alpha}{\partial u^\beta} + \dfrac{\partial f^\alpha}{\partial \psi^r} \dfrac{\partial \psi^r}{\partial u^\beta} \right| \neq 0$. Reverting to our first notation $x^\alpha = u^\alpha$ and applying (1.4) this condition on the transformation $x^A \to \bar{x}^A$ takes the form $| X_\beta f^\alpha | = | \partial_\beta f^\alpha + L^r{}_\beta \partial_r f^\alpha | \neq 0$ for points (x^A) of $S^q(c^i)$.

It has just been shown how every coordinate transformation $x^A \rightarrow \bar{x}^A$ satisfying $| X_\beta f^\alpha | \neq 0$ induces an isomorphic non-singular parameter transformation $u^\alpha \rightarrow \bar{u}^\alpha$. Thus the coordinate scalars $\lambda_\alpha = \Lambda_Q \dfrac{\partial x^Q}{\partial u^\alpha}$ will transform as parameter vectors under a coordinate transformation $x \rightarrow \bar{x}$ and because of the covariantive parameterization $x^\alpha = u^\alpha$ these parameter vectors will have the semblance of q-component coordinate vectors. The components λ_α are $\lambda_\alpha(x^A) = \Lambda_\alpha + \Lambda_r L^r{}_\alpha$ where the x's are confined to a subspace $S^q(c^i)$. But since any point (x^A) is a point of some such subspace, the quantities λ_α constitute a q-component *factor* field of the N-component vector field Λ_A. Analogous conclusions may be drawn from the complementary system (1.2) leading to the complete decomposition $\lambda_\alpha = \Lambda_Q L^Q{}_\alpha = \Lambda_\alpha + \Lambda_r L^r{}_\alpha$ and $\lambda_i = \Lambda_Q L^Q{}_i = \Lambda_i + \Lambda_\rho L^\rho{}_i$ of Λ_A into a q and $N - q$ component factor field respectively. The restriction $| X_j f^i | = | \partial_j f^i + L^\rho{}_j \partial_\rho f^i | \neq 0$ on the coordinate transformation $\bar{x}^A = f^A(x)$ analogous to $| X_\beta f^\alpha | \neq 0$ is encountered. To determine the transformation character of the factor fields we shall investigate the behavior of the basic systems of partial differential equations under a change of coordinates.

2. The Transformation Character of the Basic Systems

Let the most general coordinate transformation $x \rightarrow \bar{x}$ be sought which will preserve the solved form of the basic systems. That is, we stipulate the general laws

$$\text{(2.1)} \qquad \begin{aligned} \bar{X}_\beta f &= T_{\bar{\beta}}{}^\rho X_\rho f \qquad & \bar{X}_j f &= T_{\bar{j}}{}^r X_r f \\ P^\alpha &= \bar{P}^\rho T_{\bar{\rho}}{}^\alpha \qquad & P^i &= \bar{P}^r T_{\bar{r}}{}^i \end{aligned}$$

wherein $| T_{\bar{\beta}}{}^\alpha | \neq 0$ and $| T_{\bar{j}}{}^i | \neq 0$ so that $X_\beta f = 0$ will imply $\bar{X}_\beta f = \bar{\partial}_\beta f + \bar{L}^r{}_\beta \bar{\partial}_r f = 0$ etc. If both sides of these equations be written as linear forms in $\bar{\partial}_A f$ and $d\bar{x}^A$ and the coefficients be identified there will result

$$\text{(2.2)} \qquad \begin{aligned} \bar{L}^i{}_\rho X_\beta \bar{x}^\rho &= X_\beta \bar{x}^i \qquad & \bar{L}^\alpha{}_r X_j \bar{x}^r &= X_j \bar{x}^\alpha \\ T_{\bar{\beta}}{}^\rho X_\rho \bar{x}^\alpha &= \delta^\alpha{}_\beta \qquad & T_{\bar{j}}{}^r X_r \bar{x}^i &= \delta^i{}_j \end{aligned}$$

$$\text{(2.3)} \qquad \begin{aligned} \bar{L}^r{}_\beta \bar{X}_r x^i &= -\bar{\partial}_\beta x^i + L^i{}_\rho \bar{\partial}_\beta x^\rho \qquad & \bar{L}^\rho{}_j \bar{X}_\rho x^\alpha &= -\bar{\partial}_j x^\alpha + L^\alpha{}_r \bar{\partial}_j x^r \\ \bar{L}^r{}_\beta \bar{\partial}_r x^\alpha &= -L^\alpha{}_r \bar{\partial}_\beta x^r \qquad & \bar{L}^\rho{}_j \bar{\partial}_\rho x^i &= -L^i{}_\rho \bar{\partial}_j x^\rho. \end{aligned}$$

The first line of (2.3) reduces to the inverse of the corresponding line in (2.2) when one applies the second line of (2.3). We impose the restrictions

$$(\mathrm{R}_{2.4}) \qquad | X_\beta \bar{x}^\alpha | \neq 0 \qquad | X_j \bar{x}^i | \neq 0$$

identical with those of the preceding section on the transformations $\bar{x}^A = \bar{x}^A(x)$ so that $\bar{L}^i{}_\beta$ and $\bar{L}^\alpha{}_j$ will be determined uniquely by (2.2). The second set of (2.3) may be regarded as identities which obviously hold for the identical transformation $\bar{x}^A = x^A$. Transformations $x \rightarrow \bar{x}$ satisfying $(\mathrm{R}_{2.4})$ will be called

admissible and henceforth only such transformations will be considered. Equations (2.2) are adequate to establish the chain formulas of differentiation

$$X_\beta \bar{\bar{x}}^\alpha = \bar{X}_\rho \bar{\bar{x}}^\alpha X_\beta \bar{x}^\rho \qquad X_j \bar{\bar{x}}^i = \bar{X}_r \bar{\bar{x}}^i X_j \bar{x}^r \tag{2.5}$$

giving the conjugate relationships

$$\delta^\alpha{}_\beta = \bar{X}_\rho x^\alpha X_\beta \bar{x}^\rho \qquad \delta^i{}_j = \bar{X}_r x^i X_j \bar{x}^r \tag{2.6}$$

for the transformation $\bar{\bar{x}}^A = x^A$. Comparison with the second line of (2.2) gives $T_{\bar{\beta}}{}^\alpha = \bar{X}_\beta x^\alpha$ and $T_{\bar{j}}{}^i = \bar{X}_j x^i$ and from (2.5) the transitivity of the transformations $L \longrightarrow \bar{L}$ follows readily. The transformation equations (2.1) of the basic systems are thus

$$\begin{aligned} \bar{X}_\beta f &= X_\rho f \bar{X}_\beta x^\rho \qquad & \bar{X}_j f &= X_r f \bar{X}_j x^r \\ P^\alpha &= \bar{P}^\rho \bar{X}_\rho x^\alpha \qquad & P^i &= \bar{P}^r \bar{X}_r x^i. \end{aligned} \tag{2.7}$$

3. The Group $\Sigma \frown \Sigma^*$ of Admissible Transformations

Although the transformations $x \longrightarrow \bar{x}$ have been defined as changes of coordinates, it will be convenient in this section to regard them as point transformations in order to facilitate the description of the transitive properties of certain objects transforming under them. We shall now establish the group properties of the transformations of the set Σ: $\bar{x}^A = \sigma^A(x)$ characterized by $| X_\beta \bar{x}^\alpha | \neq 0$. We first observe that the identical transformation $\bar{x}^A = x^A$ is in Σ. Secondly, if we first transform from the point (x) to the point $(\bar{x})$ by a transformation of Σ, thereby determining at $(\bar{x})$ a unique set of functions $\bar{L}^i{}_\beta(\bar{x})$, and then transform to a third point $(\bar{\bar{x}})$ by any transformation $\bar{\bar{x}}^A = \bar{\sigma}^A(\bar{x})$ of a set which we shall call $\bar{\Sigma}$, where the elements of $\bar{\Sigma}$ are characterized by the property that $| \bar{X}_\beta \bar{\bar{x}}^\alpha | \neq 0$, it is always possible to transform directly from (x) to $(\bar{\bar{x}})$ by an element of Σ, that is, by a transformation $\bar{\bar{x}}^A = \bar{\bar{x}}^A(x)$ for which $| X_\beta \bar{\bar{x}}^\alpha | \neq 0$; for it is merely necessary to form the determinant of both sides of (2.5) and observe that $| X_\beta \bar{\bar{x}}^\alpha |$ is the product of the two non-zero determinants $| X_\beta \bar{x}^\alpha |$ and $| \bar{X}_\beta \bar{\bar{x}}^\alpha |$ and so must be likewise non-zero itself. It remains to show that if $x \longrightarrow \bar{x}$ is an element of Σ the inverse transformation $x \longleftarrow \bar{x}$ is an element of $\bar{\Sigma}$, that is, we must prove that if $| X_\beta \bar{x}^\alpha | \neq 0$ then $| \bar{X}_\beta x^\alpha | \neq 0$. But this follows immediately from (2.6), $1 = | \bar{X}_\rho x^\alpha | \cdot | X_\beta \bar{x}^\rho |$. Since it has now been proved that the most general transformation $\bar{x} \longrightarrow \bar{\bar{x}}$ with which we shall be concerned is compounded from an inverse transformation of the set Σ followed by a direct transformation of Σ we have the

Theorem 3.1: *Corresponding to any choice of the functions $L^i{}_\beta(x)$ in some arbitrary but definite coordinate system x, the subset Σ of the general group G of analytic transformations $\bar{x}^A = \bar{x}^A(x)$ for which $| X_\beta \bar{x}^\alpha | \neq 0$ when taken in conjunction with all the inverse transformations of Σ constitutes a subgroup of G.*

An analogous theorem holds for the subset Σ^* of G for which $| X_j \bar{x}^i | \neq 0$. It follows that the transformations of the intersection set $\Sigma \frown \Sigma^*$ likewise form a subgroup of G.

4. Factorization of Any Tensor

The invariance of the independence condition $| L^A{}_B | \neq 0$ will now be established by proving that after an admissible transformation $x \rightarrow \bar{x}$ characterized by $| \partial_B \bar{x}^A | \neq 0$ and $| \Delta_B \bar{x}^A | = \begin{vmatrix} X_\beta \bar{x}^\alpha & 0 \\ 0 & X_j \bar{x}^i \end{vmatrix} \neq 0$ the new determinant $| \bar{L}^A{}_B |$ will be likewise non-vanishing. The transformation laws (2.7) for the double system $X_B f = 0$ become $L^Q{}_B \partial_Q f = \bar{L}^A{}_Q \Delta_B \bar{x}^Q \bar{\partial}_A f$ and the conditions that these hold for arbitrary f are $L^Q{}_B \partial_Q \bar{x}^A = \bar{L}^A{}_Q \Delta_B \bar{x}^Q$. On forming the determinant of both sides it is apparent that $| \bar{L}^A{}_B | \neq 0$ is a consequence of $| L^A{}_B | \neq 0$, $| \partial_B \bar{x}^A | \neq 0$ and $| \Delta_B \bar{x}^A | \neq 0$ as was to be proved.

We shall call the matrix $\| L^A{}_B \|$ the *factor frame* of S^N, a name suggested by the definitions $\lambda_A = \Lambda_Q L^Q{}_A$ of the factor components λ_α and λ_i corresponding to Λ_A. Thus the left members $X_B f$ of our underlying equations are factor components of the gradient $\partial_A f$. In a new coordinate system $\bar{x}$ the factor components of an arbitrary vector $\bar{\Lambda}_B$ are defined by $\bar{\lambda}_B = \bar{\Lambda}_Q \bar{L}^Q{}_B$. For the case $\Lambda_A = \partial_A f$ we know from (2.7) that $\lambda_B = \bar{\lambda}_Q \Delta_B \bar{x}^Q$. To verify that these hold for an arbitrary vector $\bar{\Lambda}_A$ we rewrite them as $(\bar{L}^A{}_Q \Delta_B \bar{x}^Q - L^Q{}_B \partial_Q \bar{x}^A) \bar{\Lambda}_A = 0$ and note that the coefficients of $\bar{\Lambda}_A$ vanish as a consequence of the definitions of $\Delta_B \bar{x}^A$ and of the transformation laws (2.2) of the L's.

Up to this point we have considered only a factorization for a covariant vector Λ_A with factor components $\lambda_A = \Lambda_Q L^Q{}_A$ transforming according to the laws $\bar{\lambda}_\alpha = \lambda_\rho \bar{X}_\alpha x^\rho$ and $\bar{\lambda}_i = \lambda_r \bar{X}_i x^r$. Because of the condition $| L^A{}_B | \neq 0$ there exists a matrix $\| M^A{}_B \|$ conjugate to $\| L^A{}_B \|$ satisfying $M^A{}_Q L^Q{}_B = \delta^A{}_B$. As a basis for defining the factor components λ^A of a contravariant vector Λ^A we shall insist that $\lambda^Q \lambda_Q = \Lambda^Q \Lambda_Q$ hold for Λ^A and Λ_A arbitrary. This leads to the definitions $\lambda^A = M^A{}_Q \Lambda^Q$. From the four sets of relations

$$
(4.1) \qquad \begin{aligned} M^\alpha{}_\beta + M^\alpha{}_r L^r{}_\beta &= \delta^\alpha{}_\beta \qquad & M^\alpha{}_j + M^\alpha{}_\rho L^\rho{}_j &= 0 \\ M^i{}_j + M^i{}_\rho L^\rho{}_j &= \delta^i{}_j \qquad & M^i{}_\beta + M^i{}_r L^r{}_\beta &= 0 \end{aligned}
$$

resulting from $M^A{}_Q L^Q{}_B = \delta^A{}_B$ these frame components are

$$
(4.2) \qquad \lambda^\alpha = M^\alpha{}_\beta (\Lambda^\beta - L^\beta{}_r \Lambda^r) \qquad \lambda^i = M^i{}_j (\Lambda^j - L^j{}_\rho \Lambda^\rho)
$$

where the coefficients $M^\alpha{}_\beta$ and $M^i{}_j$ are defined by the conjugate relations

$$
(4.3) \qquad M^\alpha{}_\rho (\delta^\rho{}_\beta - L^\rho{}_r L^r{}_\beta) = \delta^\alpha{}_\beta \qquad M^i{}_r (\delta^r{}_j - L^r{}_\sigma L^\sigma{}_j) = \delta^i{}_j .
$$

It will now be shown how a factorization of an arbitrary covariant vector will induce a corresponding factorization for an arbitrary contravariant vector. For if we define factor components λ^A of Λ^A in two coordinate systems x and $\bar{x}$ by $\lambda^A = M^A{}_Q \Lambda^Q$ and $\bar{\lambda}^A = \bar{M}^A{}_Q \bar{\Lambda}^Q$ the two sides of $\bar{\Lambda}^A \bar{\Lambda}_A = \Lambda^Q \Lambda_Q$ will be equal respectively to the two sides of $\bar{\lambda}^A \bar{\lambda}_A = \lambda^Q \lambda_Q$, or $\bar{\lambda}^\alpha \bar{\lambda}_\alpha + \bar{\lambda}^i \bar{\lambda}_i = \lambda^\rho (\bar{\lambda}_\alpha X_\rho \bar{x}^\alpha) + \lambda^r (\bar{\lambda}_i X_r \bar{x}^i)$ by virtue of our definitions of λ_A and $\bar{\lambda}_A$. By insisting that these be identities in $\bar{\lambda}_A$ we obtain the desired laws $\bar{\lambda}^\alpha = \lambda^\rho X_\rho \bar{x}^\alpha$ and $\bar{\lambda}^i = \lambda^r X_r \bar{x}^i$. It involves but a slight extension of this demonstration to define factor com-

ponents $\lambda^A{}_B$ of a simple mixed tensor $\Lambda^A{}_B$ by $\lambda^A{}_B = M^A{}_Q\Lambda^Q{}_RL^R{}_B$ and observe that they will transform by $\bar\lambda^\alpha{}_\beta = \lambda^\rho{}_\sigma X_\rho\bar x^\alpha \bar X_\beta x^\sigma$, $\bar\lambda^i{}_j = \lambda^r{}_s X_r\bar x^i \bar X_j x^s$, $\bar\lambda^\alpha{}_j = \lambda^\rho{}_r X_\rho\bar x^\alpha \bar X_j x^r$ and $\bar\lambda^i{}_\beta = \lambda^r{}_\rho X_r\bar x^i \bar X_\beta x^\rho$. The generalization to tensors of arbitrary covariant and contravariant rank is obvious.

Let us now return to equations (4.3). Defining $t^\alpha{}_\beta$ and $t^i{}_j$ by $t^\alpha{}_\beta = \delta^\alpha{}_\beta - L^\alpha{}_rL^r{}_\beta$ and $t^i{}_j = \delta^i{}_j - L^i{}_\rho L^\rho{}_j$ the equations (4.3) are $M^\alpha{}_\rho t^\rho{}_\beta = \delta^\alpha{}_\beta$ and $M^i{}_r t^r{}_j = \delta^i{}_j$. Their form suggests that $t^\alpha{}_\beta$, $t^i{}_j$, $M^\alpha{}_\beta$, and $M^i{}_j$ are factor tensors and such is in fact the case. To establish the tensor character of $t^\alpha{}_\beta$ it is sufficient to observe that $(\bar L^\alpha{}_r\bar L^r{}_\rho)X_\beta\bar x^\rho = (L^\rho{}_sL^s{}_\beta)X_\rho\bar x^\alpha$ follows from (2.2) and (2.3). By use of the t's the transformation laws (2.2) of the L's may be written in the form

$$(4.4)\qquad \bar L^i{}_\rho X_\beta\bar x^\rho = L^r{}_\beta X_r\bar x^i + t^\rho{}_\beta\partial_\rho\bar x^i \qquad \bar L^\alpha{}_r X_j\bar x^r = L^\rho{}_j X_\rho\bar x^\alpha + t^r{}_j\partial_r\bar x^\alpha.$$

It is now apparent that corresponding to any contravariant vector Λ^A there are the two sets of factor vectors

$$\lambda^{*\alpha} = \Lambda^\alpha - L^\alpha{}_r\Lambda^r \qquad \lambda^{*i} = \Lambda^i - L^i{}_\rho\Lambda^\rho$$

$$\lambda^\alpha = M^\alpha{}_\rho\lambda^{*\rho} \qquad \lambda^i = M^i{}_r\lambda^{*r}.$$

These two sets are identical when and only when the tensors $L^\alpha{}_rL^r{}_\beta$ and $L^i{}_\rho L^\rho{}_j$ vanish.

Referring to the definitions in section 1 of the coefficients $N^A{}_B$ of the fundamental Pfaffians, it may be seen that $\| N^A{}_QL^Q{}_B \| = \begin{Vmatrix} t^\alpha{}_\beta & 0 \\ 0 & t^i{}_j \end{Vmatrix}$ and therefore $| N^A{}_Q |\cdot| L^Q{}_B | = | t^\alpha{}_\beta |\cdot| t^i{}_j |$. But since $| N^A{}_B | = | L^A{}_B |$, as may be seen by multiplying the last $N - q$ columns (index j) and the last $N - q$ rows (index i) of $| N^A{}_B |$ by -1, we have $| L^A{}_B |^2 = | t^\alpha{}_\beta |\cdot| t^i{}_j |$ and in new coordinates $\bar x$, $| \bar L^A{}_B |^2 = | \bar t^\alpha{}_\beta |\cdot| \bar t^i{}_j |$. Because the t's are simple mixed factor tensors, $| \bar t^\alpha{}_\beta | = | t^\alpha{}_\beta |$ and $| \bar t^i{}_j | = | t^i{}_j |$, and hence $| \bar L^A{}_B |^2 = | L^A{}_B |^2$. From continuity considerations under an infinitesimal transformation we finally have $| \bar L^A{}_B | = | L^A{}_B |$. Hence from $L^Q{}_B\partial_Q\bar x^A = \bar L^A{}_Q\Delta_B\bar x^Q$ there follows the relation $| \partial_B\bar x^A | = | X_\beta\bar x^\alpha |\cdot| X_j\bar x^i |$.

5. Covariant Differentiation and the Commutator Symbols

We now assume the existence in S^N of an affine connection $L^A{}_{BC}(x)$ in terms of which we define the covariant derivatives of arbitrary vectors by $\Lambda_{C,B} = \partial_B\Lambda_C - \Lambda_QL^Q{}_{BC}$ and $\Lambda^A{}_{,B} = \partial_B\Lambda^A + \Lambda^QL^A{}_{BQ}$. Similarly for the factor components of these vectors we define factor covariant derivatives by $\lambda_{C,B} = X_B\lambda_C - \lambda_Ql^Q{}_{BC}$ and $\lambda^A{}_{,B} = X_B\lambda^A + \lambda^Ql^A{}_{BQ}$, where the l's are functions to be called factor components of connection, undetermined as yet, such that these latter quantities $\lambda_{C,B}$ and $\lambda^A{}_{,B}$ will be factor components of the former tensors $\Lambda_{C,B}$ and $\Lambda^A{}_{,B}$ respectively; $\lambda_{C,B} = \Lambda_{R,Q}L^R{}_CL^Q{}_B$ and $\lambda^Q{}_{,B}L^A{}_Q = \Lambda^A{}_{,Q}L^Q{}_B$. These conditions determine

$$(5.1)\qquad l^Q{}_{BC}L^A{}_Q = L^A{}_{QR}L^Q{}_BL^R{}_C + L^Q{}_B\partial_QL^A{}_C$$

as the relations that must exist between $l^A{}_{BC}$ and $L^A{}_{BC}$. In a new coordinate system $\bar{x}$ we have components of affine connection given by $\bar{L}^Q{}_{BC}\bar{\partial}_Q x^A = L^A{}_{QR}\bar{\partial}_B x^Q \bar{\partial}_C x^R + \bar{\partial}^2_{BC} x^A$ and these induce the transformation law

$$l^Q{}_{BC}\bar{\Delta}_Q x^A = l^A{}_{QR}\bar{\Delta}_B x^Q \bar{\Delta}_C x^R + \bar{X}_B \bar{\Delta}_C x^A \tag{5.2}$$

according to which

$$\begin{aligned} l^\rho{}_{\beta\gamma}\bar{X}_\rho x^\alpha &= l^\alpha{}_{\rho\sigma}\bar{X}_\beta x^\rho \bar{X}_\gamma x^\sigma + \bar{X}_\beta \bar{X}_\gamma x^\alpha & l^\rho{}_{j\gamma}\bar{X}_\rho x^\alpha &= l^\alpha{}_{r\sigma}\bar{X}_j x^r \bar{X}_\gamma x^\sigma + \bar{X}_j \bar{X}_\gamma x^\alpha \\ l^r{}_{jk}\bar{X}_r x^i &= l^i{}_{rs}\bar{X}_j x^r \bar{X}_k x^s + \bar{X}_j \bar{X}_k x^i & l^r{}_{\beta k}\bar{X}_r x^i &= l^i{}_{\rho s}\bar{X}_\beta x^\rho \bar{X}_k x^s + \bar{X}_\beta \bar{X}_k x^i \end{aligned} \tag{5.3}$$

while $l^i{}_{\beta\gamma}$, $l^\alpha{}_{jk}$, $l^i{}_{j\gamma}$ and $l^\alpha{}_{\beta k}$ transform as factor tensors.

The skew-symmetric portion $\Omega^A{}_{BC}$ of the affine connection $L^A{}_{BC} = \Gamma^A{}_{BC} + \Omega^A{}_{BC}$ is a tensor with factor components $\omega^A{}_{BC}$ which may be computed from (5.1),

$$\tfrac{1}{2}(L^A{}_{QR} - L^A{}_{RQ})L^Q{}_B L^R{}_C = \tfrac{1}{2}(l^Q{}_{BC} - l^Q{}_{CB})L^A{}_Q - \tfrac{1}{2}I^A{}_{BC} \tag{5.4}$$

where $I^A{}_{BC} = L^Q{}_B\partial_Q L^A{}_C - L^Q{}_C\partial_Q L^A{}_B$ so that $I^\alpha{}_{\beta\gamma} = I^i{}_{jk} = 0$. The commutator $(X_B X_C)f$ is then $(X_B X_C)f = I^Q{}_{BC}\partial_Q f = M^A{}_Q I^Q{}_{BC} L^R{}_A \partial_R f = \lambda^A{}_{BC} X_A f$ where $\lambda^A{}_{BC} = M^A{}_Q I^Q{}_{BC}$. The quantities $\lambda^A{}_{BC}$ will be called *commutator symbols* for the operator X_B and in terms of them we have by (5.4), $\Omega^A{}_{QR}L^Q{}_B L^R{}_C = \tfrac{1}{2}[(l^Q{}_{BC} - l^Q{}_{CB}) - \lambda^Q{}_{BC}]L^A{}_Q$, and hence the factor components $\omega^A{}_{BC}$ of $\Omega^A{}_{BC}$ are

$$\omega^A{}_{BC} = \tfrac{1}{2}[(l^A{}_{BC} - l^A{}_{CB}) - \lambda^A{}_{BC}]. \tag{5.5}$$

6. Integrability Conditions and Normal Coordinates

The system $X_\beta f = 0$ will be completely integrable if and only if the commutators $(X_\beta X_\gamma)f = \lambda^\rho{}_{\beta\gamma}X_\rho f + \lambda^r{}_{\beta\gamma}X_r f$ are linear combinations of $X_\beta f$. Hence $\lambda^i{}_{\beta\gamma} = M^i{}_r I^r{}_{\beta\gamma} = 0$ are the conditions for integrability. Because of the tensor character of $M^i{}_j$ and the circumstance that $| M^i{}_j | \neq 0$ we may write $I^i{}_{\beta\gamma} = 0$ in place of $\lambda^i{}_{\beta\gamma} = 0$ and for this reason we shall call $I^i{}_{\beta\gamma}$ an integrability tensor. When $I^i{}_{\beta\gamma} = 0$ there then exist $N - q$ independent solutions $f^i(x)$ which represent ∞^{N-q} subspaces $S^q(c^j)$ of q dimensions when equated to arbitrary constants c^i.

Let us now consider the more general non-holonomic case wherein $I^i{}_{\beta\gamma} \neq 0$ and therefore $X_\beta f = 0$ are not completely integrable. In this instance it becomes more convenient to discuss the Pfaffian system $dx^i - L^i{}_\rho\, dx^\rho = 0$ corresponding to $X_\beta f = 0$ rather than these latter equations themselves. These equations determine an elemental q-spread of directions dx^A at points of S^N and in the holonomic case the totality of these elemental q-spreads envelops the ∞^{N-q} subspaces $S^q(c^i)$. For the sake of generality this non-holonomic viewpoint will be retained henceforth and the holonomic theory will become the special case characterized by $I^i{}_{\beta\gamma} = 0$. Similar remarks apply to the Pfaffian system $dx^\alpha - L^\alpha{}_r\, dx^r = 0$ corresponding to the complementary equations $X_i f = 0$ whose integrability conditions are $I^\alpha{}_{jk} = 0$.

The question naturally arises as to what conditions may serve to assure the

existence of an admissible coordinate system y^A for which one set of the L's, say $L^\alpha{}_j(y)$, vanishes. This is answered by the following

THEOREM 6.1: *The factor tensor conditions*

$$I^\alpha{}_{jk} = X_j L^\alpha{}_k - X_k L^\alpha{}_j = 0 \tag{6.1}$$

$$L^\alpha{}_r L^r{}_\beta = 0 \tag{6.2}$$

are necessary and sufficient for the existence of an admissible coordinate transformation from general coordinates x^A to normal coordinates y^A for which $L^\alpha{}_j(y) = 0$.

PROOF: Since $I^\alpha{}_{jk}$ and $L^\alpha{}_r L^r{}_\beta$ are factor tensors, if there is to exist a coordinate system y for which $L^\alpha{}_j(y) = 0$ it is certain that the conditions of the theorem must necessarily hold. To show their sufficiency we have in virtue of (6.1) that the system $X_j y = \partial_j y + L^\rho{}_j \partial_\rho y = 0$ admits q independent principal solutions $y^\alpha(x)$ satisfying the initial conditions $y^\alpha(x_\beta, x_0^k) = x^\alpha$. Let $y^i(x)$ be any $N - q$ functions independent of y^α for which $| X_j y^i | \neq 0$. We assert that $y^A = y^A(x)$ constitutes a non-singular transformation to coordinates y for which $L^\alpha{}_j(y) = 0$. From the transformation law $\bar{L}^\alpha{}_r X_j \bar{x}^r = X_j \bar{x}^\alpha$ it follows on identifying $\bar{x}^A$ with y^A that $L^\alpha{}_j(y) = 0$. Furthermore the transformation $x \to y$ is admissible, for $| X_\beta y^\alpha | = | \partial_\beta y^\alpha + L^r{}_\beta \partial_r y |$ has the initial value $| \delta^\alpha{}_\beta - L^\alpha{}_r L^r{}_\beta |_{x^k = x_0^k} = | \delta^\alpha{}_\beta |$ by (6.2). Finally from $| \partial_B y^A | = | X_\beta y^\alpha | \cdot | X_j y^i |$ we see that $| \partial_B y^A | \neq 0$ so that the transformation $x \to y$ is non-singular. A similar proof holds for the complementary

THEOREM 6.2: *The conditions $I^i{}_{\beta\gamma} = X_\beta L^i{}_\gamma - X_\gamma L^i{}_\beta = 0$ and $L^i{}_\rho L^\rho{}_j = 0$ are necessary and sufficient for the existence of coordinates y^A for which $L^i{}_\beta(y) = 0$.*

7. AN AFFINE CONNECTION FOR S^N

By means of the equations (5.1) the factor components $l^A{}_{BC}$ of connection are uniquely determined by the components $L^A{}_{BC}$ and conversely these latter components are uniquely determined by a choice of $l^A{}_{BC}$. In this section we shall adopt this converse viewpoint and shall define the factor components $l^A{}_{BC}$ in the simplest way at our disposal. We first recall from section 5 that four sets of the connection components $l^A{}_{BC}$ are factor tensors. We now restrict these to be zero

$$l^i{}_{\beta\gamma} = 0 \qquad l^\alpha{}_{jk} = 0 \qquad l^i{}_{j\gamma} = 0 \qquad l^\alpha{}_{\beta k} = 0. \tag{R$_{7.1}$}$$

From $\omega^A{}_{BC} = \frac{1}{2}[(l^A{}_{BC} - l^A{}_{CB}) - \lambda^A{}_{BC}]$ as given by (5.5) it follows as a result of (R$_{7.1}$) that the quantities $\omega^\alpha{}_{j\gamma} = \frac{1}{2}(l^\alpha{}_{j\gamma} - \lambda^\alpha{}_{j\gamma})$ and $\omega^i{}_{\beta k} = \frac{1}{2}(l^i{}_{\beta k} - \lambda^i{}_{\beta k})$ are factor tensors. We therefore make the additional restrictions

$$l^\alpha{}_{j\gamma} = \lambda^\alpha{}_{j\gamma} \qquad l^i{}_{\beta k} = \lambda^i{}_{\beta k} \tag{R$_{7.2}$}$$

which leave us with but two undefined sets of components $l^\alpha{}_{\beta\gamma}$ and $l^i{}_{jk}$. Thus an entire affine connection for S^N is determined by (R$_{7.1}$), (R$_{7.2}$) and definitions of $l^\alpha{}_{\beta\gamma}$ and $l^i{}_{jk}$.

The restrictions just imposed on the factor components of affine connection simplify the formulas of covariant differentiation as follows:

$$
(7.3)\qquad
\begin{aligned}
\lambda_{\gamma,\beta} &= X_\beta\lambda_\gamma - \lambda_\rho l^\rho{}_{\beta\gamma} & \lambda_{k,j} &= X_j\lambda_k - \lambda_r l^r{}_{jk}\\
\lambda_{\gamma,j} &= X_j\lambda_\gamma - \lambda_\rho\lambda^\rho{}_{j\gamma} & \lambda_{k,\beta} &= X_\beta\lambda_k - \lambda_r\lambda^r{}_{\beta k}\\
\lambda^i{}_{,\beta} &= X_\beta\lambda^i + \lambda^r\lambda^i{}_{\beta r} & \lambda^\alpha{}_{,j} &= X_j\lambda^\alpha + \lambda^\rho\lambda^\alpha{}_{j\rho}\\
\lambda^i{}_{,j} &= X_j\lambda^i + \lambda^r l^i{}_{jr} & \lambda^\alpha{}_{,\beta} &= X_\beta\lambda^\alpha + \lambda^\rho l^\alpha{}_{\beta\rho}.
\end{aligned}
$$

We define parallel displacement of a contravariant vector Λ^A along a curve $x^A = x^A(s)$, where s is an affine parameter, by the equations $\Lambda^A{}_{,Q}\dfrac{dx^Q}{ds} = 0$. For a curve lying in a subspace S^q, holonomic or non-holonomic, the tangent vector $T^A = \dfrac{dx^A}{ds}$ with factor components $t^A = M^A{}_Q\dfrac{dx^Q}{ds}$ satisfies the Pfaffian system $t^i = M^i{}_r\left(\dfrac{dx^r}{ds} - L^i{}_\rho\dfrac{dx^\rho}{ds}\right) = 0$. From $\dfrac{dx^A}{ds} = L^A{}_Q t^Q$ it then follows that $t^\alpha = \dfrac{dx^\alpha}{ds}$. The factor components of $\Lambda^A{}_{,Q}\dfrac{dx^Q}{ds} = 0$ may be written $L^S{}_Q\partial_S\lambda^A M^Q{}_R\dfrac{dx^R}{ds} + l^A{}_{RS}t^R\lambda^S = 0$ or $\dfrac{d\lambda^A}{ds} + l^A{}_{RS}t^R\lambda^S = 0$ so that for parallel displacement along a curve in S^q we have

$$
(7.4)\qquad \frac{d\lambda^\alpha}{ds} + l^\alpha{}_{\rho\sigma}\frac{dx^\rho}{ds}\lambda^\sigma = 0 \qquad \frac{d\lambda^i}{ds} + \lambda^i{}_{\rho r}\frac{dx^\rho}{ds}\lambda^r = 0 \qquad \frac{dx^i}{ds} - L^i{}_\rho\frac{dx^\rho}{ds} = 0.
$$

For parallel displacement of a covariant vector Λ_A in S^q the corresponding system $\Lambda_{A,Q}\dfrac{dx^Q}{ds} = 0$ has the factor components

$$
(7.5)\qquad \frac{d\lambda_\alpha}{ds} - \lambda_\rho l^\rho{}_{\sigma\alpha}\frac{dx^\sigma}{ds} = 0 \qquad \frac{d\lambda^i}{ds} - \lambda_r\lambda^r{}_{\rho i}\frac{dx^\rho}{ds} = 0 \qquad \frac{dx^i}{ds} - L^i{}_\rho\frac{dx^\rho}{ds} = 0.
$$

A path of S^q is defined as a curve whose tangent vector $\lambda^\alpha = \dfrac{dx^\alpha}{ds}$, $\lambda^i = 0$, undergoes parallel displacement with respect to the curve. From (7.4) the equations of a path of S^q are

$$
(7.6)\qquad \frac{d^2x^\alpha}{ds^2} + l^\alpha{}_{\rho\sigma}\frac{dx^\rho}{ds}\frac{dx^\sigma}{ds} = 0 \qquad \frac{dx^i}{ds} - L^i{}_\rho\frac{dx^\rho}{ds} = 0.
$$

8. The Riemannian S^N

In this section we shall consider only the completely holonomic case for which $\lambda^A{}_{\beta\gamma} = \lambda^A{}_{jk} = 0$ and shall assume that S^N is Riemannian with a metric tensor G_{AB}. Representing the Christoffel symbols by $\Gamma^A{}_{BC} = \frac{1}{2}G^{AQ}(\partial_B G_{CQ} + \partial_C G_{BQ} - \partial_Q G_{BC})$ and their factor components by $\left\{ {A \atop BC} \right\}$ these latter quantities will be given

in terms of $\Gamma^A{}_{BC}$ and $L^A{}_B$ by relations of the type (5.1). We desire the explicit form of $\left\{ {A \atop BC} \right\}$ in terms of the factor components $g_{AB} = G_{QR}L^Q{}_A L_{RB}$ of the metric tensor and for this purpose shall make use of the circumstance that $G_{AB,C} = 0$ and hence its frame components $g_{AB,C} = X_C g_{AB} - g_{QB}\left\{ {Q \atop CA} \right\} - g_{AQ}\left\{ {Q \atop CB} \right\}$ must likewise vanish. Substituting into $\frac{1}{2}(g_{AB,C} + g_{AC,B} - g_{BC,A}) = 0$ and using the relations $\omega^A{}_{BC} = \frac{1}{2}\left[\left(\left\{ {A \atop BC} \right\} - \left\{ {A \atop CB} \right\}\right) - \lambda^A{}_{BC}\right] = 0$ expressing the vanishing of the factor components of the skew-symmetric portion of $\Gamma^A{}_{BC}$ we obtain after minor rearrangements $\frac{1}{2}g_{AQ}\left(\left\{ {Q \atop BC} \right\} + \left\{ {Q \atop CB} \right\}\right) = [BC, A] - \frac{1}{2}(g_{BQ}\lambda^Q{}_{CA} + g_{CQ}\lambda^Q{}_{BA})$, where $[BC, A] = \frac{1}{2}(X_B g_{CA} + X_C g_{BA} - X_A g_{BC})$. Combining these with $\frac{1}{2}g_{AQ}\left(\left\{ {Q \atop BC} \right\} - \left\{ {Q \atop CB} \right\}\right) = \frac{1}{2}g_{AQ}\lambda^Q{}_{BC}$ we arrive at the desired expressions

$$(8.1) \qquad g_{AQ}\left\{ {Q \atop BC} \right\} = [BC, A] - \tfrac{1}{2}(g_{BQ}\lambda^Q{}_{CA} + g_{CQ}\lambda^Q{}_{BA} - g_{AQ}\lambda^Q{}_{BC}).$$

We now ask if it is possible to restrict the metric tensor g_{AB} so that the quantities $\left\{ {A \atop BC} \right\}$ will satisfy the conditions ($R_{7.1}$) and ($R_{7.2}$) imposed in section 7 on the general affine connection $l^A{}_{BC}$. It may be verified that the conditions

$$(R_{8.2}) \qquad \lambda^i{}_{\beta\gamma} = 0 \qquad \lambda^\alpha{}_{jk} = 0 \qquad g_{i\alpha} = 0$$

in conjunction with ($R_{7.1}$) and ($R_{7.2}$) are sufficient to assure compatibility, for on substituting these values into (8.1) we are left with but two sets of non-vanishing relations $g_{\alpha\rho}\left\{ {\rho \atop \beta\gamma} \right\} = [\beta\gamma, \alpha]$ and $g_{ir}\left\{ {r \atop jk} \right\} = [jk, i]$ defining the components $\left\{ {\alpha \atop \beta\gamma} \right\}$ and $\left\{ {i \atop jk} \right\}$ which together with ($R_{7.1}$) and ($R_{7.2}$) define the complete factor components $\left\{ {A \atop BC} \right\}$ of the Christoffel symbols $\Gamma^A{}_{BC}$. Under these conditions there exist holonomic coordinates z^A for which $L^A{}_B(z) = \delta^A{}_B$ and the metric tensor $G_{AB}(z) = g_{AB}(z)$ is in the product form $\left\| \begin{matrix} G_{\alpha\beta}(z^\gamma) & 0 \\ 0 & G_{ij}(z^k) \end{matrix} \right\|$ as follows from $g_{\alpha\beta,k} = 0$ and $g_{ij,\gamma} = 0$ when written in normal form.

9. Contact Transformations

In the study of contact transformations a change in notation is desirable. Here we are concerned with $N = 2q$ variables x^A commonly denoted by x^i and p_i. We adapt the general factorization notation of the preceding sections to the contact group in $2q$ variables by choosing the two subranges to be $1, \cdots, q$ and $q + 1, \cdots, 2q$. Vector components will be written in the form (covariant)

$\| \Lambda_A \| = \|\Lambda_i, \Lambda^i\|$, (contravariant) $\| \Lambda^A \| = \| \Lambda^i, \Lambda_i \|$. This convention is in need of an additional notation to distinguish between covariant and contravariant quantities with indices in the same position but we shall deal only with a few tensors whose transformation character is clearly understood so that no ambiguity will arise. A similar scheme will be used for non-tensor quantities; for example we shall write $L^i{}_{q+j} = L^{ij}$ and $L^{q+i}{}_j = L_{ij}$.

The full contact transformation group in $2q$ variables x^i and p_i may be defined as the totality of transformations $X^i = X^i(x, p)$, $P_i = P_i(x, p)$ for which $P_r\,dX^r - p_r\,dx^r$ regarded as an expression in the $2q$ variables x^i and p_i and their differentials is the complete differential of some function of the x's and p's, that is,

$$P_r(x, p)\,dX^r(x, p) - p_r\,dx^r = dW(x, p). \tag{9.1}$$

More generally, a Pfaffian form $\varphi_Q(u)\,du^Q$ in $2q$ variables u^A will be said to be *preserved* modulo a complete differential under the transformation $u \to \bar{u}$ if the relation $(\varphi_Q(\bar{u})\partial_B\bar{u}^Q - \varphi_B(u))\,du^B = dW(u)$ holds, the necessary and sufficient conditions on the coefficients of du^B being $\varphi_{QR}(\bar{u})\partial_A\bar{u}^Q\partial_B\bar{u}^R = \varphi_{AB}(u)$, where $\varphi_{AB} = \partial_A\varphi_B - \partial_B\varphi_A$. From the general theory of Pfaffian forms the condition $|\varphi_{AB}| \neq 0$ is necessary and sufficient for the existence of a transformation $u \to x$ sending the Pfaffian $\varphi_Q\,du^Q$ into a canonical form $p_r\,dx^r$. A contact transformation $x^A \to \bar{x}^A$ may therefore be described as the most general transformation preserving the canonical form modulo a complete differential. For the canonical Pfaffian $\Pi_Q\,dx^Q = p_r\,dx^r$ the matrix of $\Pi_{AB} = \partial_A\Pi_B - \partial_B\Pi_A$ is simply

$$\| \Pi_{AB} \| = \begin{Vmatrix} \Pi_{ij} & \Pi_i{}^j \\ \Pi^i{}_j & \Pi^{ij} \end{Vmatrix} = \begin{Vmatrix} 0 & -\delta_i^j \\ \delta^i{}_j & 0 \end{Vmatrix} \tag{9.2}$$

and the tensor transformation law $\Pi_{AQ}\bar{\partial}_B x^Q = \Pi_{QB}\partial_A\bar{x}^Q$ of its elements contains in concise form the familiar equations of definition of the contact group

$$-\bar{\partial}_j p_i = \partial_i\bar{p}_j \qquad -\bar{\partial}^j x^i = \partial^i\bar{x}^j \qquad \bar{\partial}^j p_i = \partial_i\bar{x}^j \tag{9.3}$$

as was first recognized by Eisenhart and Knebelman (1).

When the function $W(x, p)$ of (9.1) vanishes identically it may be shown that $P_i(x, p)$ and $X^i(x, p)$ are homogeneous in the p's of degrees one and zero respectively. For this reason transformations $x^A \to \bar{x}^A$ for which $\bar{p}_r\,d\bar{x}^r = p_r\,dx^r$ are called homogeneous contact transformations and they form a subgroup of the full contact group. By introducing two new variables, x^{q+1} and $p_{q+1} = -1$ and extending the transformation $x^A \to \bar{x}^A$ to

$$\begin{aligned} \bar{x}^i &= \bar{x}^i(x, p) & \bar{p}_i &= \bar{p}_i(x^j, p_k) \\ \bar{x}^{q+1} &= x^{q+1} + W(x^i, p_j) & \bar{p}_{q+1} &= p_{q+1} = -1 \end{aligned}$$

(9.1) becomes $\sum_1^{q+1} \bar{p}_\rho\,d\bar{x}^\rho = \sum_1^{q+1} p_\rho\,dx^\rho$. Since any contact transformation in $2q$ variables may be written in this way as a homogeneous contact transformation in $2q + 2$ variables we shall restrict our discussion henceforth to homogeneous transformations in $2q$ variables.

The coefficients $\| \Pi_A \| = \| p_i , 0 \|$ of the Pfaffian $p_r \, dx^r$ form a vector whose transformation law $\Pi_Q \partial_A \bar{x}^Q = \Pi_A$ gives the relations

$$P_r \partial_i X^r = p_i \qquad P_r \partial^i X^r = 0 \tag{9.4}$$

which are equivalent to (9.3) plus the homogeneity relations $p_r \partial^r X^i = 0$ and $p_r \partial^r P_i = P_i$ and either set fully characterizes the homogeneous contact group. We shall need the factor components $\pi_A = \Pi_Q L^Q{}_A$ and $\pi_{AB} = \Pi_{QR} L^Q{}_A L^R{}_B$ for future reference. They are

$$\| \pi_A \| = \| p_i, \, p_r L^{ri} \|$$
$$\| \pi_{AB} \| = \begin{Vmatrix} \pi_{ij} & \pi_i{}^j \\ \pi^i{}_j & \pi^{ij} \end{Vmatrix} = \begin{Vmatrix} -(L_{ij} - L_{ji}) & -(\delta^j{}_i - L_{ri} L^{rj}) \\ (\delta^i{}_j - L_{rj} L^{ri}) & (L^{ij} - L^{ji}) \end{Vmatrix}. \tag{9.5}$$

10. Normal Coordinates for the Contact Frame

THEOREM 10.1: *The factor tensor conditions*

$$I^{ijk} = X^j L^{ik} - X^k L^{ij} = 0 \quad \pi^{ij} = L^{ij} - L^{ji} = 0 \quad \pi^i = p_r L^{ri} = 0 \tag{$R_{10.1}$}$$

are necessary and sufficient for the existence of a homogeneous contact transformation from general coordinates x^A to normal coordinates y^A for which $L^{ij}(y) = 0$.

PROOF: The necessity of these conditions is obvious from their linearity in L^{ij} and their tensor character. To prove their sufficiency we notice from the transformation equations $\bar{L}^{jr} X^k \bar{p}_r = X^k \bar{x}^j$ of the frame components L^{jk} that if we identify $\bar{x}^A$ with the desired normal coordinates y^A we must choose $Y^i(x, p)$ as solutions of the system $X^k Y = \partial^k Y + L^{rk} \partial_r Y = 0$. The conditions $I^{ijk} = 0$ assure complete integrability of these equations so that we may find q functionally independent solutions $Y^i(x, p)$ satisfying

$$\partial^k Y^i + L^{rk} \partial_r Y^i = 0, \tag{10.2}$$

and by the theory of section 1 for equations of this solved type the solutions Y^i satisfy $| \partial_k Y^i | \neq 0$. Contraction of equations (10.2) with p_k shows that the solutions Y^i are homogeneous in the p's of degree zero. Because $| \partial_j Y^i | \neq 0$ we may define q functions $Q_i(x, p)$ as the unique solutions of $Q_r \partial_j Y^r = p_j$ which will be homogeneous in the p's of degree one. By the contraction $Q_r(\partial^k Y^r + L^{sk} \partial_s Y^r) = 0$ we find $Q_r \partial^k Y^r = 0$ and hence $y^i = Y^i(x, p)$ and $q^i = Q_i(x, p)$ satisfy the conditions (9.4) for a homogeneous contact transformation.

We next ask for the most general contact transformation preserving normal coordinates. By identifying x^A with y^A and $\bar{x}^A$ with $\bar{y}^A$ in $\bar{L}^{ir} X^j \bar{p}_r = L^{rj} X_r \bar{x}^i + t_r{}^j \partial^r \bar{x}^i$ as given by (4.4), it follows immediately that any transformation from normal coordinates y for which $L^{ij}(y) = 0$ to a system $\bar{y}$ for which also $\bar{L}^{ij}(\bar{y}) = 0$ must satisfy $\dfrac{\partial \bar{y}^i}{\partial q_j} = 0$ and therefore must be an extended point transformation. Hence the

THEOREM 10.2: *The most general homogeneous contact transformation $y^A \to \bar{y}^A$ preserving normal coordinates for which $L^{ij} = 0$ is an extended point transformation.*

11. Restrictions on the Contact Frame

We shall impose henceforth the restrictions ($R_{10.1}$) assuring the existence of normal coordinates. If we define a *union* to be any subspace of S^{2q} within which an arbitrary displacement dx^A at a point (x^A) satisfies $p_r\,dx^r = 0$ it may be shown that such a subspace may be regarded as a point locus in the point space of coordinates x^i together with all its tangent hyperplane elements with coordinates p_i. Thus the maximum dimensionality of a union is $q - 1$. In terms of unions the restrictions ($R_{10.1}$) require that any q independent solutions of the system $X^i f = 0$ when equated to arbitrary constants shall constitute a union of maximum dimensionality $q - 1$, for in normal coordinates the system becomes $\frac{\partial f}{\partial q_i} = 0$ and its solutions y^i determine the ∞^q unions $y^i = y^i_0$ consisting of the point (y^i_0) in the point space Y^q regarded as the envelope of its ∞^{q-1} hyperplane elements q_i.

By operating on the transformation laws (2.2) with the invariant operator $p_r\partial^r$ it may be shown that the quantities $h_{jk} = L_{jk} - p_r\partial^r L_{jk}$ and $h^{jk} = L^{jk} + p_r\partial^r L^{jk}$ are covariant factor tensors, and since h^{jk} vanishes in normal coordinates, $L^{jk} + p_r\partial^r L^{jk} = 0$. As a second set of restrictions on the contact frame we introduce

$$(R_{11.1})\qquad -\pi_{jk} = L_{jk} - L_{kj} = 0 \qquad h_{jk} = L_{jk} - p_r\partial^r L_{jk} = 0.$$

We next collect for future reference some of the more useful consequences of the restrictions of this section. From equations (2.3) we have $\bar{L}_{rk}\bar{\partial}^r x^i = -L^{ir}\bar{\partial}_j p_r$ and hence $X_j\bar{x}^i = \bar{X}^i p_j$. The transformation laws of factor vectors simplify to $\bar{\lambda}_j = \lambda_r \bar{X}_j x^r$ and $\bar{\lambda}^j = \lambda^r X_r \bar{x}^j$ for both the covariant vector $\| \lambda_A \| = \| \lambda_i, \lambda^i \|$ and the contravariant vector $\| \lambda^A \| = \| \lambda^i, \lambda_i \|$ so that an upper index of covariance has contravariant character and a lower index of contravariance has covariant character. Under extended point transformations $\bar{y}^i = \bar{y}^i(y^j)$, $\bar{q}_i = q_r\bar{\partial}_i y^r$, these laws reduce to the familiar form $\bar{\lambda}_j(\bar{y}) = \lambda_r(y)\bar{\partial}_j y^r$, $\bar{\lambda}^j(\bar{y}) = \lambda^r(y)\partial_r \bar{y}^j$. The two tensors $L^i{}_\rho L^\rho{}_j$ and $L^\alpha{}_r L^r{}_\beta$ of section 4 coincide to give just one vanishing tensor $L^{ir}L_{rj} = 0$ and the two mixed tensors $t^\alpha{}_\beta$ and $t^i{}_j$ reduce to the single Kroencker delta $t^i{}_j = \delta^i{}_j$. The transformation laws (4.4) of the L's are now

$$(11.2)\qquad \bar{L}_{ir}\partial_j\bar{x}^r = L_{rj}\bar{\partial}_i x^r + \partial_j\bar{p}_i \qquad \bar{L}^{ir}\bar{\partial}_r x^j = L^{rj}\partial_r\bar{x}^i + \partial^j\bar{x}^i$$

and from (9.5) the factor tensors π_A and π_{AB} have the same components as Π_A and Π_{AB}, $\| \pi_A \| = \| p_i, 0 \|$, $\| \pi_{AB} \| = \begin{Vmatrix} 0 & -\delta_i{}^j \\ \delta^i{}_j & 0 \end{Vmatrix}$. The matrix $\| M^A{}_B \|$ conjugate to $\| L^A{}_B \|$ is $\begin{Vmatrix} M^i{}_j & M^{ij} \\ M_{ij} & M_i{}^j \end{Vmatrix} = \begin{Vmatrix} \delta^i{}_j & -L^{ij} \\ -L_{ij} & \delta_i{}^j \end{Vmatrix}$. In normal coordinates y^A the commutator symbols $\lambda^A{}_{BC}$ have the values $\lambda^i{}_{jk}(y) = \lambda^{ijk}(y) = \lambda^{ij}{}_k(y) = 0$, $\lambda_{ijk}(y) = Y_j L_{ik} - Y_k L_{ij}$, $\lambda_{ij}{}^k(y) = -\partial^k L_{ij}$.

12. A Complete Determination of an Affine Connection for Contact Transformations

We carry over into this section the restrictions ($R_{7.1}$) and ($R_{7.2}$) of the general theory of section 7. In our present notation these restrictions are

$$(R_{12.1}) \qquad l_{ijk} = l^{ijk} = l_i{}^j{}_k = l^i{}_j{}^k = 0, \qquad l^{ij}{}_k = \lambda^{ij}{}_k, \qquad l_{ij}{}^k = \lambda_{ij}{}^k,$$

and so a complete determination of factor components of affine connection will result when we have selected the two sets of components $l^i{}_{jk}$ and $l_i{}^{jk}$. This choice will be made by insisting that the covariant derivative of the factor tensor π_{AB} be zero. From (7.3) the only conditions are $X_j\pi_k{}^i - \pi_r{}^i l^r{}_{jk} - \pi_k{}^r\lambda_{rj}{}^i = 0$ and $X^j\pi_i{}^k - \pi_r{}^k\lambda^{rj}{}_i - \pi_k{}^r l_r{}^{jk} = 0$, which reduce by the values of the π's given in section 11 to

$$(R_{12.2}) \qquad l^i{}_{jk} + \lambda_{kj}{}^i = 0 \qquad l_i{}^{jk} + \lambda^{kj}{}_i = 0.$$

These restrictions define the components $l^i{}_{jk}$ and $l_i{}^{jk}$ so that a complete affine connection is given by ($R_{12.1}$) and ($R_{12.2}$).

On expanding $\lambda_{kj}{}^i$ and $\lambda^{kj}{}_i$ as defined in section 5 we find $\lambda_{kj}{}^i = -\partial^i L_{kj} + L^{ir}I_{kjr}$ and $\lambda^{kj}{}_i = -\partial_i L^{kj} + L_{ir}I^{kjr} = -\partial_i L^{kj}$ so that $l^i{}_{jk} = \partial^i L_{kj} - L^{ir}I_{kjr}$ and $l_i{}^{jk} = \partial_i L^{jk}$. By contraction with p_i we find $p_r l^r{}_{jk} = L_{jk}$ which will be interpreted in section (14). The formulas of covariant differentiation coincide in pairs to give

$$\lambda_{k,j} = X_j\lambda_k - \lambda_r l^r{}_{jk} \qquad \lambda^{k,j} = X^j\lambda^k + \lambda^r\lambda^{kj}{}_r$$

$$\lambda_{k,}{}^j = X^j\lambda_k - \lambda_r\lambda^{rj}{}_k \qquad \lambda^k{}_{,j} = X_j\lambda^k + \lambda^r l^k{}_{jr},$$

where the vector λ is either covariant with components $(\lambda_A) = (\lambda_i, \lambda^i)$ or contravariant with components $(\lambda^A) = (\lambda^i, \lambda_i)$.

13. Holonomic Coordinates for the Contact Frame

If we operate on (5.1), $X_D L^A{}_B = l^Q{}_{DB}L^A{}_Q - L^A{}_{QR}L^Q{}_D L^R{}_B$ with $X_C = L^Q{}_C\partial_Q$ and then form $(X_C X_D)L^A{}_B = \lambda^Q{}_{CD}X_Q L^A{}_B$ we obtain $l^Q{}_{BCD}L^A{}_Q = L^A{}_{QRS}L^Q{}_B L^R{}_C L^S{}_D$, where $L^A{}_{BCD}$ is the curvature tensor $L^A{}_{BCD} = \partial_C L^A{}_{DB} - \partial_D L^A{}_{CB} + L^A{}_{CQ}L^Q{}_{DB} - L^A{}_{DQ}L^Q{}_{CB}$ and the quantities $l^A{}_{BCD} = X_C l^A{}_{DB} - X_D l^A{}_{CB} + l^A{}_{CQ}l^Q{}_{DB} - l^A{}_{DQ}l^Q{}_{CB} - \lambda^Q{}_{CD}l^A{}_{QB}$ are its factor components. Under the restrictions of sections 11 and 12 the only non-vanishing factor components of the contact curvature tensor are

$$l^i{}_{jkl} = -l_j{}^i{}_{kl} = X_l\lambda_{jk}{}^i - X_k\lambda_{jl}{}^i + \lambda_{rk}{}^i\lambda_{jl}{}^r - \lambda_{rl}{}^i\lambda_{jk}{}^r + \lambda^r{}_{kl}\lambda_{jr}{}^i - \lambda_{rkl}\lambda^{ir}{}_j$$

$$l^i{}_{jk}{}^l = \quad l_j{}^{il}{}_k = X_k\lambda^{il}{}_j + X^l\lambda_{jk}{}^i + \lambda^{il}{}_r\lambda_{jk}{}^r - \lambda_{rk}{}^i\lambda^{rl}{}_j - \lambda^{rl}{}_k\lambda_{jr}{}^i - \lambda_{rk}{}^l\lambda^{ir}{}_j$$

and they reduce in normal coordinates y^A to

$$l^i{}_{jkl}(y) = \partial_k(\partial^i L_{lj}) - \partial_l(\partial^i L_{kj}) + \partial^i L_{kr}\partial^r L_{lj} - \partial^i L_{lr}\partial^r L_{kj}$$

$$+ L_{rk}\partial^{2ri}L_{lj} - L_{rl}\partial^{2ri}L_{kj} \tag{13.1}$$

$$l^i{}_{jk}{}^l(y) = -\partial^{2li}L_{jk}.$$

We next seek the conditions for the existence of holonomic coordinates $(z^A) = (z^i, r_i)$ for which simultaneously $L^{jk}(z) = 0$ and $L_{jk}(z) = 0$.

THEOREM 13.1: *The conditions*

$$I^{ijk} = X^jL^{ik} - X^kL^{ij} = 0 \qquad \pi^{ij} = L^{ij} - L^{ji} = 0 \quad \pi^i = p_rL^{ri} = 0 \tag{13.2}$$

$$I_{ijk} = X_jL_{ik} - X_kL_{ij} = 0 \quad -\pi_{ij} = L_{ij} - L_{ji} = 0 \quad l^i{}_{jk}{}^l = 0 \tag{13.3}$$

are necessary and sufficient for the existence of a homogeneous contact transformation from general coordinates x^A to holonomic coordinates z^A for which $L^A{}_B(z) = \delta^A{}_B$.

PROOF: From theorem 10.1 the first set of conditions is necessary and sufficient for the existence of normal coordinates y^A for which $L^{ij}(y) = 0$. In this coordinate system the left members of the last of (13.3) reduce to (13.1) so that if there is to exist a coordinate system z for which $L_{jk}(z) = 0$ it is at once obvious that the conditions (13.3) must necessarily hold. The sufficiency of the conditions of the theorem may be established by first applying theorem 10.1 which asserts the sufficiency of conditions (13.1) for the existence of normal coordinates y^A. We then apply theorem (10.2) which states that the most general homogeneous contact transformation $y \to z$ preserving $L^{ij} = 0$ is an extended point transformation,

$$T\colon z^i = z^i(y^j) \qquad r_i = q_r\frac{\partial y^r}{\partial z^i} \qquad T^{-1}\colon y^i = y^i(z^j) \qquad q_i = r_s\frac{\partial z^s}{\partial y^i}.$$

It remains to be shown that the conditions (13.3) are sufficient for the existence of q independent functions $z^i(y^j)$ such that T will constitute a transformation to coordinates z^A for which $L_{jk}(z) = 0$ as well as $L^{jk}(z) = 0$. To accomplish this we note that the transformation law $L_{jr}(y)\left(\frac{\partial y^r}{\partial z^k} + L_{sk}(z)\frac{\partial y^r}{\partial r_s}\right) = \frac{\partial q_j}{\partial z^k} + L_{sk}(z)\frac{\partial q_j}{\partial r_s}$ reduces to $L_{jr}(y)\frac{\partial y^r}{\partial z^k} = q_i\frac{\partial y^i}{\partial z^s}\frac{\partial}{\partial y^r}\left(\frac{\partial z^s}{\partial y^j}\right)\frac{\partial y^r}{\partial z^k}$ for an extended point transformation T, and since $\left|\frac{\partial y^r}{\partial z^k}\right| \neq 0$, $L_{jk}(y) = q_i\frac{\partial y^i}{\partial z^s}\frac{\partial}{\partial y^k}\left(\frac{\partial z^s}{\partial y^j}\right)$. Differentiating with respect to q_i and solving for $\frac{\partial}{\partial y^k}\left(\frac{\partial z^i}{\partial y^j}\right)$ we obtain $\frac{\partial}{y^k}\left(\frac{\partial z^i}{\partial y^j}\right) = \frac{\partial z^i}{\partial y^r}\partial^rL_{jk}$ and thus arrive at the linear system

$$\partial_jz^i = Z_j{}^i(y^k) \tag{13.4}$$

$$\partial_kZ_j{}^i = Z_r{}^i\partial^rL_{jk} \tag{13.5}$$

in $q(q+1)$ unknowns z^i and $Z_j{}^i$ to be determined as functions of the q variables y^i. Since the quantities $Z_j{}^i$ may be functions of only y^i it is necessary that $\partial^{2il}L_{jk}(y) = 0$ and these are satisfied by the last of (13.3). The integrability

conditions of (13.4) are found by expressing the conditions that $Z_r^{\ i}(\partial^r L_{jk} - \partial^r L_{kj})$ shall vanish for all $Z_j^{\ i}$, namely $\partial^i(L_{jk} - L_{kj}) = 0$, and these are satisfied by the middle set of (13.3). Finally the integrability conditions of (13.5) are $\partial_l(\partial^i L_{jk}) - \partial_k(\partial^i L_{jl}) + \partial^i L_{rl}\partial^r L_{jk} - \partial^i L_{rk}\partial^r L_{jk} = 0$ and these are satisfied by the first of (13.3) as may be seen by differentiating in normal coordinates with respect to q_i. Finally from the general theory of completely integrable systems of the type (13.4) and (13.5) there exist solutions $Z_j^{\ i}$ for which $| Z_j^{\ i} | = | \partial_j z^i | \neq 0$. This completes the proof of the theorem.

14. Introduction of a Metric Tensor

The trajectories of a one parameter homogeneous contact group are solutions $x^A(s)$ of the system

$$(14.1) \qquad \frac{dx^i}{ds} = \partial^i C \qquad \frac{\partial p_i}{ds} = -\partial_i C$$

whose vector character is obvious when written in the form $\Pi_{AQ}\dfrac{dx^Q}{ds} = \partial_A C$. $C(x, p)$ is necessarily homogeneous of degree one in the p's and hence we may take $C = 1$ providing we exclude the singular locus $C = 0$ from our discussion. Since C is an integral of (14.1) the condition $C = 1$ will hold at all points of a trajectory if it holds at one point.

We now choose the scalar C to be a solution of $X_j f = 0$,

$$(\mathrm{R}_{14.2}) \qquad \partial_j C + L_{rj}\partial^r C = 0.$$

Then the second set of (14.1) becomes $\dfrac{dp_i}{ds} = L_{ri}\dfrac{dx^r}{ds}$. These equations have been discussed in section 7 where it was seen that they express the vanishing of λ_i as defined by $\dfrac{dx^A}{ds} = L^A{}_Q\lambda^Q$ so that these latter reduce to $\lambda^i = \dfrac{dx^i}{ds}$ and $\lambda_i = \dfrac{dp_i}{ds} - L_{ir}\dfrac{dx^r}{ds} = 0$. Thus the displacement dx^A is restricted to the non-holonomic subspace X^q of S^{2q}, where X^q is the point space of points (x^i) and is characterized as a non-holonomic subspace of S^{2q} by the arbitrariness of the differentials dx^i. Furthermore it may be seen from (7.5) that the equations of parallel displacement of the vector $(\pi_A) = (p_i, 0)$ are $\dfrac{dp_i}{ds} - p_r\Gamma^r_{si}\dfrac{dx^s}{ds} = 0$, but from the last of section 12 these become $\lambda_i = 0$ so that the trajectories under consideration have the property that the vector π_A undergoes parallel displacement along them.

We shall now introduce a metric tensor $G_{AB}(x)$, whose factor components g_{AB} will be of the restricted form $\| g_{AB} \| = \begin{Vmatrix} g_{ij} & 0 \\ 0 & g^{ij} \end{Vmatrix}$ where $g_{ir}g^{rj} = \delta_i^{\ j}$. We define g^{ij} in terms of the scalar $H = \frac{1}{2}C^2$, where C is the scalar satisfying $X_j C = 0$, by $g^{ij}(x, p) = (X^i H)^{,k}$ reducing in normal coordinates to $g^{ij}(y, q) = \partial^{2ij}H(y, q)$. From these definitions the components $g^{ij}(x, p)$ are homogeneous

of degree zero in the p's. There is a theorem (4) which states that a necessary and sufficient condition for the Hessian determinant $| \partial^{2ij} H |$ to be non-singular is that $| \partial^{2ij} C |$ be of maximum rank $q - 1$. Restricting our discussion to those solutions C of $X_i f = 0$ for which the rank of $| \partial^{2ij} C |$ is $q - 1$ we have the two fundamental requirements, $g^{ij} = g^{ji}$ and $| g^{ij} | \neq 0$, of a metric tensor fulfilled. On doubly contracting the tensor g^{ij} with the covariant factor vector $\pi_i = p_i$ we obtain the invariant $g^{rs} p_r p_s$ which in normal coordinates is $\partial^{2rs} H q_r q_s = 2H$ and hence in all coordinates $H = \frac{1}{2} g^{rs} p_r p_s$. The system of trajectories (14.1) which may equally well be written $\dfrac{dx^i}{ds} = \partial^i H$ and $\dfrac{dp_i}{ds} = -\partial_i H$ is now

$$\frac{dx^i}{ds} = g^{ir} p_r \qquad \frac{dp_i}{ds} = -\tfrac{1}{2} \partial_i g^{rs} p_r p_s, \tag{14.3}$$

or in alternative form

$$p_i = g_{ir} \frac{dx^r}{ds} \qquad \frac{dp_i}{ds} = L_{ir} \frac{dx^r}{ds}. \tag{14.4}$$

We next seek additional restrictions on the frame $L^A{}_B$ and the metric g^{ij} which will identify the trajectories (14.3) with the geodesics of the subspace X^q. From equations (7.6) defining the paths of the point space X^q we arrive at the equations of the geodesics by substituting the factor components of the Christoffel symbols in place of the affine connections components. Thus the geodesics are given by

$$\frac{d^2 x^i}{ds^2} + \begin{Bmatrix} i \\ rs \end{Bmatrix} \frac{dx^r}{ds} \frac{dx^s}{ds} = 0 \qquad \frac{dp_i}{ds} - L_{ir} \frac{dx^r}{ds} = 0 \tag{14.5}$$

and it is this system that is to be identified with (14.3).

15. Additional Restrictions on the Contact Frame

We now investigate the possibility of restricting g^{ij} to be free of the q's in normal coordinates, $\partial^k g^{ij}(y^A) = 0$, as expressed by $g^{ij,k} = 0$. This will necessitate that $H(y^A) = \frac{1}{2} g^{rs}(y^i) q_r q_s$ be a solution of $\partial_i H + L_{ri} \partial^r H = 0$. The conditions are

$$\tfrac{1}{2} \partial_i g^{rs} q_r q_s = -L_{ir} g^{rs} q_s \tag{15.1}$$

and by repeated differentiation we derive the necessary conditions

$$\partial_i g^{jk} = -(\partial^k L_{ir} g^{rj} + \partial^j L_{ir} g^{rk}) - \partial^{2jk} L_{ir} g^{rs} q_s .$$

If we introduce the curvature tensor restrictions

$$l^i{}_{jk}{}^l = 0 \tag{R$_{15.2}$}$$

reducing in normal coordinates to $\partial^{2il} L_{jk}(y) = 0$ we are left with

$$\partial_i g^{jk} = -(\partial^k L_{ir} g^{rj} + \partial^j L_{ir} g^{rk}) \tag{15.3}$$

which may be solved for the quantities $\partial^i L_{jk}$ as Christoffel symbols formed from $G_{ij}(y^A) = g_{ij}(y^k)$, $\partial^i L_{jk}(y) = \Gamma^i{}_{jk}(y^l)$, and hence

$$L_{ij}(y^l, q_m) = q_r \Gamma^r{}_{jk}(y^l). \tag{15.4}$$

Conversely, if the expressions (15.3) resulting from (15.4) by differentiation are substituted into (15.1) it may be verified that these latter equations are satisfied.

The restrictions just imposed on g^{ij} and L_{jk} are sufficient to bring the geodesics (14.5) into coincidence with the trajectories (14.3), for differentiation of the first of (14.3) in normal coordinates gives

$$\frac{d^2 y^i}{ds^2} - \Gamma^i{}_{rs}(y^j)\frac{dy^r}{ds}\frac{dy^s}{ds} = 0 \qquad \frac{dq_i}{ds} = L_{ir}(y^A)\frac{dy^r}{ds}. \tag{15.5}$$

Under the present restrictions the equations (8.1) define $\left\{ {i \atop jk} \right\}$ by $\left\{ {i \atop jk} \right\} = \frac{1}{2}g^{ir}[(X_j g_{kr} + X_k g_{jr} - X_r g_{jk}) - (g_{js}\lambda^s{}_{kr} + g_{ks}\lambda^s{}_{jr} - g_{js}\lambda^s{}_{jk})]$ which reduce in normal coordinates to the familiar Christoffel symbols, $\{{}^i{}_{jk}(y)\} = \Gamma^i{}_{jk}(y^l)$. Thus the normal form of the equations (14.5) of the geodesics is likewise given by (15.5) and because of the vector character of (14.3) and (14.5) this completes the proof of their equivalence. It is interesting to note that from the first of (14.4) and the relations $p_r\lambda^r{}_{jk} = 0$ resulting from the definitions $\lambda^i{}_{jk} = -L^{ir}I_{rjk}$ of section 5, the first of (14.5) is given by $\frac{d^2x^i}{ds} + g^{it}[\frac{1}{2}(X_r g_{st} + X_s g_{rt} - X_t g_{rs})]\frac{dx^r}{ds}\frac{dx^s}{ds} = 0$. The tensor $l^i{}_{jk} - \left\{ {i \atop jk} \right\}$ reduces in normal coordinates to $\partial^i L_{jk} - \Gamma^i{}_{jk}(y^l) = 0$ and hence in all coordinates $l^i{}_{jk} = \left\{ {i \atop jk} \right\}$. The paths of X^q are therefore identical with the geodesics. We shall regard the affine connection of S^{2q} to be that of section 12 where the components $l^i{}_{jk}$ as there defined now satisfy the additional relations $l^i{}_{jk} = \left\{ {i \atop jk} \right\}$. Because of ($\mathrm{R}_{15.2}$) the curvature tensor $l^A{}_{BCD}$ has but one set of non-vanishing components $l^i{}_{jkl} = -l_j{}^i{}_{kl}$ reducing in normal coordinates to the Riemann curvature tensor $R^i{}_{jkl}(y^m) = \partial_k\Gamma^i{}_{lj} - \partial_l\Gamma^i{}_{kj} + \Gamma^i{}_{kr}\Gamma^r{}_{lj} - \Gamma^i{}_{lr}\Gamma^r{}_{kj}$ formed from the Christoffel symbols $\Gamma^i{}_{jk}(y^l)$.

The results of this section may be summarized by the statement that the three ostensibly distinct systems of invariantive curves consisting of

1) the trajectories of $\Pi_{AQ}\frac{dx^Q}{ds} = \partial_A C$ where $X_j C = 0$,

2) the geodesics of the point space X^q with respect to the metric tensor $g^{ij}(x, p) = \frac{1}{2}(X^i C^2)^{\cdot j}$,

3) the paths of X^q relative to the connection $l^i{}_{jk} = \partial^i L_{kj} - L^{ir}I_{kjr}$ have been combined into just one system.

16. Doubly Homogeneous Contact Transformations

Let Σ be a projective space of $N - 1$ dimensions, S an arbitrary hypersurface of Σ parameterized by the equations $x^i = \varphi^i(u^\alpha)$, $i = 1, 2, \cdots, N$, $\alpha =$

$1, 2, \cdots, N-2$, and P an arbitrary point of S. The conditions on a set of hyperplane coordinates p_i that they represent the tangent hyperplane at P may be found by insisting that the plane p_i cut an arbitrary curve $u^\alpha = u^\alpha(t)$ in two coincident points at P whose parameter we choose to be $t = 0$. This means that $p_r x^r(t) = at^2 + \cdots$ when written as a power series in t. Hence the conditions for $t = 0$ are $p_r x^r = 0$ and $p_r \dfrac{\partial x^r}{\partial u^\alpha}\dfrac{du^\alpha}{dt} = 0$. But since these must hold for an arbitrary curve $u^\alpha(t)$ the conditions on the point-hyperplane element $(x^i(u), p_i(u))$ that it represent a point $x^i(u)$ incident with the plane $p_i(u)$ which in turn is tangent to the hypersurface S at x^i are $p_r x^r = 0$ and $p_r \dfrac{\partial x^r}{\partial u^\alpha} = 0$. But dually we may regard S as a locus of hyperplanes $p_i(u)$ and ask for the conditions on a point x^i that it be a "tangent point" of S. By analogous reasoning these conditions are $p_r x^r = 0$ and $x^r \dfrac{\partial p_r}{\partial u^\alpha} = 0$. An $N-2$ spread of point-hyperplane elements $(x^i(u), p_i(u))$ will be said to constitute a *union* after the terminology of Lie if the conditions $p_r x^r = 0$, $p_r \dfrac{\partial x^r}{\partial u^\alpha} = 0$, $x^r \dfrac{\partial p_r}{\partial u^\alpha} = 0$ are satisfied as identities in the arbitrary parameters u^α.

Suppose now that the union of elements $x^A(u)$ viewed in Σ either as a point locus $x^i = x^i(u)$ or as a hyperplane locus $p_i(u)$ is subjected to a transformation $\bar{x}^A = \bar{x}^A(x^B)$. If the new elements as functions of the same parameters u^α are to form a union with the point locus $\bar{x}^i = \bar{x}^i(x^B(u))$ and hyperplane locus $\bar{p}_i = \bar{p}_i(x^B(u))$ as a consequence of being the transform of a union we must have $\bar{p}_r\bar{x}_r = \lambda(x, p)p_r x^r$, $\bar{p}_r \dfrac{\partial \bar{x}^r}{\partial u^\alpha} = \mu(x, p)p_r \dfrac{\partial x^r}{\partial u^\alpha}$, $\bar{x}^r \dfrac{\partial \bar{p}}{\partial u^\alpha} = \nu(x, p)x^r \dfrac{\partial p_r}{\partial u^\alpha}$ for an arbitrary parameterization u^α. Expanding the last two conditions we have $\bar{p}_r\partial_i\bar{x}^r = p_i$, $\bar{p}_r\partial^i\bar{x}^r = 0$, $\bar{x}^r\partial^i\bar{p}_r = \nu x^i$, $\bar{x}^r\partial_i\bar{p}_r = 0$, and differentiation of the first with respect to x^i and p_i gives $(\lambda - \mu)p_i + \partial_i\lambda p_r x^r = 0$ and $(\lambda - \nu)x^i + \partial^i\lambda p_r x^r = 0$. But these latter conditions must hold identically both when $p_r x^r = 0$ and when $p_r x^r \neq 0$ so that $\partial_A\lambda = 0$ and $\lambda = \mu = \nu =$ constant. There will be two distinct types of such transformations depending on whether $\lambda > 0$ or $\lambda < 0$. In all generality we may take $\lambda = 1$ in the first case and $\lambda = -1$ in the second giving the two types

$$(\mathrm{I_a})\quad \text{(Proper contact transformation)}\quad \begin{matrix} \bar{p}_r\partial_i\bar{x}^r = p_i \\ \bar{p}_r\partial^i\bar{x}^r = 0 \end{matrix} \quad \bar{p}_r\bar{x}^r = p_r x^r \quad \begin{matrix} \bar{x}^r\partial^i\bar{p}_r = x^i \\ \bar{x}^r\partial_i\bar{p}_r = 0 \end{matrix}$$

$$(\mathrm{II_a})\quad \text{(Improper contact transformation)}\quad \begin{matrix} \bar{p}_r\partial_i\bar{x}^r = -p_i \\ \bar{p}_r\partial^i\bar{x}^r = 0 \end{matrix} \quad \bar{p}_r\bar{x}^r = -p_r x^r \quad \begin{matrix} \bar{x}^r\partial^i\bar{p}_r = -x^i \\ \bar{x}^r\partial_i\bar{p}_r = 0. \end{matrix}$$

Let us next consider an element transformation sending the point locus $x^i = x^i(u)$ into the hyperplane locus $\bar{p}_i = \bar{p}_i(x^B(u))$, and dually, the same locus when regarded as a hyperplane locus $p_i = p_i(u)$ into the same barred locus

regarded as a point locus $\bar{x}^i = \bar{x}^i(x^B(u))$. The conditions that the transformed elements $\bar{x}^A(u)$ constitute a union if the original elements $x^A(u)$ form a union are $\bar{p}_r\bar{x}^r = \lambda' p_r x^r$, $\bar{p}_r \dfrac{\partial \bar{x}^r}{\partial u^\alpha} = \mu' x^r \dfrac{\partial p_r}{\partial u^\alpha}$, $\bar{x}^r \dfrac{\partial \bar{p}_r}{\partial u^\alpha} = \nu' p_r \dfrac{\partial x^r}{\partial u^\alpha}$, and again it follows that $\lambda' = \mu' = \nu' =$ constant so that the transformations under discussion fall into two classifications:

$$(\mathrm{I_b})\quad \text{(Proper contact correlation)}\quad \begin{matrix} \bar{p}_r\partial_i\bar{x}^r = 0 \\ \bar{p}_r\partial^i\bar{x}^r = x^i \end{matrix} \quad \bar{p}_r\bar{x}^r = p_r x^r \quad \begin{matrix} \bar{x}^r\partial^i\bar{p}_r = 0 \\ \bar{x}^r\partial_i\bar{p}_r = p_i \end{matrix}$$

$$(\mathrm{II_b})\quad \text{(Improper contact correlation)}\quad \begin{matrix} \bar{p}_r\partial_i\bar{x}^r = 0 \\ \bar{p}_r\partial^i\bar{x}^r = -x^i \end{matrix} \quad \bar{p}_r\bar{x}^r = -p_r x^r \quad \begin{matrix} \bar{x}^r\partial^i\bar{p}_r = 0 \\ \bar{x}^r\partial_i\bar{p}_r = -p_i . \end{matrix}$$

By differentiating either the equations $\mathrm{I_a}$ or $\mathrm{II_b}$ one may verify the bracket relationships

$$(\text{for } \mathrm{I_a} \text{ and } \mathrm{II_b})\qquad \begin{aligned} \{x^i, x^j\} &= \partial_i\bar{p}_r\partial_j\bar{x}^r - \partial_j\bar{p}_r\partial_i\bar{x}^r = 0, \\ \{p_i, p_j\} &= \partial^i\bar{p}_r\partial^j\bar{x}^r - \partial^j\bar{p}_r\partial^i\bar{x}^r = 0 \\ \{p_i, x^j\} &= \partial^i\bar{p}_r\partial_j\bar{x}^r - \partial_j\bar{p}_r\partial^i\bar{x}^r = \delta^i{}_j . \end{aligned}$$

Similarly the transformations $\mathrm{II_a}$ or $\mathrm{I_b}$ yield

$$(\text{for } \mathrm{II_a} \text{ and } \mathrm{I_b})\qquad \{x^i, x^j\} = 0, \qquad \{p_i, p_j\} = 0, \qquad \{p_i, x^j\} = -\delta^i{}_j .$$

Using a device originated by L. P. Eisenhart and M. S. Knebelman (1) we employ the bracket relations to identify the expressions

$$\begin{aligned} dx^i &= \bar{\partial}_j x^i\, d\bar{x}^j + \bar{\partial}^j x^i\, d\bar{p}_j \qquad & dp_i &= \bar{\partial}_j p_i\, d\bar{x}^j + \bar{\partial}^j p_i\, d\bar{p}_j \\ dx^i &= \partial^i \bar{p}_j\, d\bar{x}^j - \partial^i\bar{x}^j\, d\bar{p}_j \qquad & dp_i &= -\partial_i\bar{p}_j\, d\bar{x}^j + \partial_i\bar{x}^j\, d\bar{p}_j \end{aligned}$$

with the subsequent results

$$(\text{for } \mathrm{I_a} \text{ and } \mathrm{II_b})\qquad \bar{\partial}_j x^i = \partial^i\bar{p}_j , \qquad \bar{\partial}^j x^j = -\partial^i\bar{x}^j, \qquad \bar{\partial}_j p_i = -\partial_i\bar{p}_j .$$

In like manner we find

$$(\text{for } \mathrm{II_a} \text{ and } \mathrm{I_b})\qquad \bar{\partial}_j x^i = -\partial^i\bar{p}_j , \qquad \bar{\partial}^j x^i = \partial^i\bar{x}^j, \qquad \bar{\partial}_j p_i = \partial_i\bar{p}_j .$$

These serve to establish the homogeneity properties

$$(\text{for } \mathrm{I_a} \text{ and } \mathrm{II_a})\quad \begin{aligned} \bar{x}^i(cx, kp) &= c\bar{x}^i(x, p) \\ \bar{p}_i(cx, kp) &= k\bar{p}_i(x; p) \end{aligned} \qquad (\text{for } \mathrm{I_b} \text{ and } \mathrm{II_b})\quad \begin{aligned} \bar{x}^i(cx, kp) &= k\bar{x}^i(x, p) \\ \bar{p}_i(cx, kp) &= c\bar{p}_i(x, p) \end{aligned}$$

so that the transformations may be written

$$(\mathrm{I_a} \text{ and } \mathrm{II_a})\quad \bar{x}^i = x^r\partial_r\bar{x}^i, \quad \bar{p}_i = p_r\partial^r\bar{p}_i \quad (\mathrm{I_b} \text{ and } \mathrm{II_b})\quad \bar{x}^i = p_r\partial^r\bar{x}^i, \quad \bar{p}_i = x^r\partial_r\bar{p}_i .$$

It may be verified that the four elements $\mathrm{I_a}$, $\mathrm{II_a}$, $\mathrm{I_b}$, $\mathrm{II_b}$ form a group with the identity $\mathrm{I_a}$ which is isomorphic with Klein's four-group.

17. A Complete Determination of a Factor Frame from an Affine Connection in a Flat Element Space

We consider in this section a flat element space in which there exist normal coordinates y^A for which the affine connection components $L^A{}_{BC}(y) = 0$. Since $\dot{\Omega}^A{}_{BC} = 0$ the connection is symmetric and will be written as $\Gamma^A{}_{BC}$. Confining ourselves to the doubly homogeneous contact group, we purpose to define a factor frame $L^A{}_B$ in terms of $\Gamma^A{}_{BC}$ and the element vector (x^A). To this end we first observe that the invariant numerical tensor Π_{CA} has a vanishing covariant derivative in normal coordinates, and hence in all coordinates $\Pi_{QA}\Gamma^Q{}_{BC} + \Pi_{CQ}\Gamma^Q{}_{BA} = 0$, or

$$(17.1)\qquad \begin{array}{lll} \Gamma_{ijk} = \Gamma_{kji} & \Gamma_{ij}{}^k = -\Gamma^k{}_{ji} & \Gamma^i{}_j{}^k = \Gamma^k{}_j{}^i \\ \Gamma^{ijk} = \Gamma^{kji} & \Gamma_i{}^{jk} = -\Gamma^{kj}{}_i & \Gamma_i{}^j{}_k = \Gamma_k{}^j{}_i. \end{array}$$

Since the covariant derivative of the vector x^A satisfies $x^A{}_{,B} - \delta^A{}_B = 0$ in normal coordinates we have the zero tensor $x^Q\Gamma^A{}_{BQ}$ whose expansion gives

$$(17.2)\quad p_r\Gamma^r{}_{ji} = x^r\Gamma_{rji}, \quad p_r\Gamma^{rji} = x^r\Gamma_r{}^{ji}, \quad p_r\Gamma^r{}_j{}^i = x^r\Gamma_{rj}{}^i, \quad p_r\Gamma^{rj}{}_i = x^r\Gamma_r{}^j{}_i$$

when reference is made to (17.1). We define the factor frame L_{ij} and L^{ij} by $L_{ij} = \Gamma_{ij} = p_r\Gamma^r{}_{ij}$ and $L^{ij} = \Gamma^{ij} = p_r\Gamma^{rij}$ so that $\Gamma_{ij} = \Gamma_{ji}$ and $\Gamma^{ij} = \Gamma^{ji}$ and shall demonstrate that the quantities Γ_{ij} and Γ^{ij} satisfy the conditions for a factor frame. First we note that the normal coordinates y^A for which $\Gamma^A{}_{BC}(y) = 0$ are likewise normal for $L^A{}_B$ since $L^A{}_B(y) = \delta^A{}_B$. Recalling the results of section 3, the transformation group now under discussion is the intersection group $\Sigma \frown \Sigma^*$ where Σ and Σ^* are characterized as all transformations from normal coordinates y^A to general coordinates x^A for which $| Y_i x^j | = \left|\frac{\partial x^j}{\partial y^i}\right| \neq 0$ and $| Y^j p_i | = \left|\frac{\partial p_i}{\partial q_j}\right| \neq 0$. But since $\frac{\partial x^j}{\partial y^i} = \frac{\partial p_i}{\partial q_j}$, Σ and Σ^* coincide. Furthermore, from $\delta^i{}_j = X_j y^r Y_r x^i = Y^r p_j Y_r x^i = \partial^r p_j \partial_r x^i = \partial^i q_r \partial_j y^r$ and from the bracket relationship $\partial^i q_r \partial_j y^r - \partial_j q_r \partial^i y^r = \delta^i{}_j$ it follows that the transformations $y \rightarrow x$ must satisfy

$$(17.3)\qquad \partial_j q_r \partial^i y^r = 0.$$

If in the transformation law

$$(17.4)\qquad \Gamma^A{}_{BC} = \bar{\Gamma}^Q{}_{RS}\bar{\partial}_Q x^A \partial_B \bar{x}^R \partial_C \bar{x}^S + \bar{\partial}_Q x^A \partial^2{}_{BC}\bar{x}^Q$$

we identify $\bar{x}^A$ with normal coordinates y^A, we have $\Gamma^A{}_{BC}(x) = \partial_Q x^A \partial^2{}_{BC} y^Q$, and on contracting with x^i, $x^r\Gamma_{rj}{}^k = x^r(\partial_s p_r \partial^2{}_j{}^k y^s + \partial^s p_r \partial^2{}_j{}^k q_s) = y^s\partial^2{}_j{}^k q_s$. But on differentiating $y^r\partial_s p_j = 0$, $y^s\partial^k(\partial_s p_j) + \partial^k y^s \partial_s p_j = -(y^s\partial^2{}_j{}^k q_s + \partial^k y^s \partial_j q_s) = 0$, and therefore

$$(17.5)\qquad x^r\Gamma_{rj}{}^k = x^r\Gamma_r{}^k{}_j = 0.$$

To obtain the transformation laws of the frame components sum (17.4) in the form $\bar{\Gamma}^Q{}_{BR}\bar{\partial}_Q x^A \partial_C \bar{x}^R = \Gamma^A{}_{QC}\bar{\partial}_B x^Q + \bar{\partial}^2{}_{BQ} x^A \partial_C \bar{x}^Q$ with the covariant vector $(x_A) =$

(p_i, x^j), $\bar{x}_Q \bar{\Gamma}^Q{}_{BR} \partial_C \bar{x}^R = x_Q \Gamma^Q{}_{RC} \bar{\partial}_B x^R + x_Q \partial_C (\bar{\partial}_B x^Q)$, and expand with reference to (17.2) and (17.5). The resulting equations fall into the three sets $\bar{L}_{jr} \bar{\partial}^r x^k = -L^{kr} \bar{\partial}_j p_r$, $\bar{L}_{jr} \partial_k \bar{x}^r = L_{rk} \bar{\partial}_j x^r + \partial_k \bar{p}_j$, $\bar{L}^{jr} \bar{\partial}_r x^k = L^{rk} \partial_r \bar{x}^j + \partial^k \bar{x}^j$ agreeing with the second line of (2.3) and with the form (11.2) of the first line of (2.2). This completes the identification of the quantities Γ_{ij} and Γ^{ij} as components of a factor frame.

18. A Real Decomposition Under the Complex Group

Let $(x^A) = (x^k, x^{\dot k})$ be $2q$ real variables in terms of which q complex variables z^k are defined by $z^k = x^k + i x^{\dot k}$. If z^k be subjected to an analytic transformation $\bar{z}^k = \bar{z}^k(z^j)$ the $2q$ real variables $(\bar{x}^A) = (\bar{x}^k, \bar{x}^{\dot k})$ defined by $\bar{z}^k = \bar{x}^k + i\bar{x}^{\dot k}$ will satisfy the conditions

$$(18.1) \qquad \partial_j \bar{x}^k = \partial_{\dot j} \bar{x}^{\dot k} \qquad \partial_{\dot j} \bar{x}^k = -\partial_j \bar{x}^{\dot k}.$$

These generalized Cauchy-Riemann equations fully characterize the complex group $\bar{x}^A = \bar{x}^A(x)$ induced on the $2q$ real variables x^A by the group of all analytic transformations on the q complex variables z^k. The conditions (18.1) may be written more concisely as the transformation law $E^Q{}_B \partial_Q \bar{x}^A = E^A{}_Q \bar{\partial}_B \bar{x}^Q$ of a simple mixed tensor $E^A{}_B$ with the components $\left\| \begin{matrix} E^i{}_j & E^i{}_{\dot j} \\ E^{\dot i}{}_j & E^{\dot i}{}_{\dot j} \end{matrix} \right\| = \left\| \begin{matrix} 0 & -\delta^i{}_j \\ \delta^i{}_j & 0 \end{matrix} \right\|$ in all coordinate systems.

There is a marked similarity between the complex and contact transformation groups which will become increasingly evident as we proceed. Just as $L^{jk} = 0$ served to characterize extended point transformations as a subgroup of the contact group, so here from the laws $\bar{L}^k{}_r(\partial_j \bar{x}^r + L^{\dot s}{}_j \partial_{\dot s} \bar{x}^r) = \partial_j \bar{x}^{\dot k} + L^{\dot s}{}_j \partial_{\dot s} \bar{x}^{\dot k}$ and $\bar{L}^{\dot i}{}_{\dot r}(\partial_{\dot k} \bar{x}^{\dot r} + L^s{}_{\dot k} \partial_s \bar{x}^{\dot r}) = \partial_{\dot k} \bar{x}^i + L^s{}_{\dot k} \partial_s \bar{x}^i$ it may be seen on identifying x^A with a system of coordinates y^A for which

$$(\mathrm{R}_{18.2}) \qquad L^i{}_{\dot k}(y) = 0 \qquad L^{\dot k}{}_j(y) = 0$$

and $\bar{x}^A$ with a new set of coordinates $\bar{y}^A$ with the same properties that these completely characterize the subgroup of the complex group consisting of all separated transformations of the form $\bar{y}^k = \varphi^k(y^r)$ and $\bar{y}^{\dot k} = \varphi^{\dot k}(y^{\dot r})$. We accordingly define the L's by $(\mathrm{R}_{18.2})$ in some arbitrary but definite coordinate system y^A which will be called normal. As immediate consequences of these definitions we have $L^i{}_{\dot r} L^{\dot r}{}_k = 0$.

$$L^{\dot i}{}_r L^r{}_{\dot j} = 0, \qquad L^{\dot i}{}_j + L^i{}_{\dot j} = 0, \qquad X_j \bar{x}^{\dot i} = X_{\dot j} \bar{x}^i,$$

$$\| \epsilon^A{}_B \| = \| M^A{}_Q E^Q{}_R L^R{}_B \| = \left\| \begin{matrix} \epsilon^i{}_j & \epsilon^i{}_{\dot j} \\ \epsilon^{\dot i}{}_j & {}^{\dot i}{}_{\dot j} \end{matrix} \right\| = \left\| \begin{matrix} 0 & -\delta^i{}_j \\ \delta^i{}_j & 0 \end{matrix} \right\|.$$

We complete the choice of affine connection made in section 7 by stipulating that the covariant derivatives of the tensor $\epsilon^A{}_B$ shall vanish, the conditions being

$$\epsilon^{\dot i}{}_{k,j} = X_j \epsilon^{\dot i}{}_k - \epsilon^{\dot i}{}_r \Gamma^r{}_{jk} + \epsilon^{\dot r}{}_k \lambda^{\dot i}{}_{j\dot r} = 0 \qquad \epsilon^{\dot i}{}_{k,\dot j} = X_{\dot j} \epsilon^{\dot i}{}_k - \epsilon^{\dot i}{}_r \lambda^r{}_{\dot j k} + \epsilon^{\dot r}{}_k l^{\dot i}{}_{\dot j \dot r} = 0.$$

We define $l^{i}{}_{jk}$ and $l^{\dot{i}}{}_{j\dot{k}}$ as the solutions

$$(\mathrm{R}_{18.3}) \qquad l^{i}{}_{jk} = \lambda^{\dot{i}}{}_{j\dot{k}} \qquad l\,{}_{j\dot{k}} = \lambda^{i}{}_{jk}$$

of these equations. Because all components of the curvature tensor vanish for our present choice of connection the spaces S^{2q} is flat. From section 3 the subgroups Σ and Σ^* of the full complex group corresponding to our choice of frame consist of all transformations from normal coordinates y^A to general coordinates x^A satisfying

$$\left|\frac{\partial x^i}{\partial y^j} + L^{\dot{r}}{}_j(y)\,\frac{\partial x^i}{\partial y^{\dot{r}}}\right| = \left|\frac{\partial x^i}{\partial y^j}\right| \neq 0 \quad \text{and} \quad \left|\frac{\partial x^{\dot{i}}}{\partial y^{\dot{j}}} + L^{r}{}_{\dot{j}}(y)\,\frac{\partial x}{\partial y^{r}}\right| = \left|\frac{\partial x^i}{\partial y^j}\right| \neq 0$$

so that $\Sigma = \Sigma^*$.

19. A Complex Decomposition Under the Complex Group

The transformation equations of the factor frame are satisfied by the quantities $L^A{}_B$ with the components $\begin{Vmatrix} L^i{}_j & L^i{}_{\dot{j}} \\ L^{\dot{i}}{}_j & L^{\dot{i}}{}_{\dot{j}} \end{Vmatrix} = \begin{Vmatrix} \delta^i{}_j & -i\delta^i{}_j \\ -i\delta^{\dot{i}}{}_j & \delta^{\dot{i}}{}_{\dot{j}} \end{Vmatrix}$ in all coordinate systems, for they become $-i\partial_j\bar{x}^i - \partial_{\dot{j}}\bar{x}^{\dot{i}} = \partial_j\bar{x}^{\dot{i}} - i\partial_{\dot{j}}\bar{x}^{\dot{i}}$ and $-i\partial_{\dot{j}}\bar{x}^i - \partial_j\bar{x}^i = \partial_{\dot{j}}\bar{x}^i - i\partial_j\bar{x}^i$ which hold in virtue of the equations of definition of the complex group. The reciprocal matrix $\| M^A{}_B \|$ is given by $\begin{Vmatrix} M^i{}_j & M^i{}_{\dot{j}} \\ M^{\dot{i}}{}_j & M^{\dot{i}}{}_{\dot{j}} \end{Vmatrix} = \begin{Vmatrix} \frac{1}{2}\delta^i{}_j & \frac{i}{2}\delta^i{}_j \\ \frac{i}{2}\delta^i{}_j & \frac{1}{2}\delta^i{}_j \end{Vmatrix}.$

The equations

$$T: \begin{aligned} z^i &= x^i + ix^{\dot{i}} \\ z^{\dot{i}} &= x^i - ix^{\dot{i}} \end{aligned} \qquad T^{-1}: \begin{aligned} x^i &= \frac{1}{2}(z^i + z^{\dot{i}}) \\ x^{\dot{i}} &= \frac{i}{2}(z^{\dot{i}} - z^i) \end{aligned}$$

represent a transformation from the $2q$ real variables x^A to the q complex variables z^i and their conjugates $z^{\dot{i}}$. The vector components Z and factor components λ of a real covariant vector Λ_A and of a real contravariant vector Λ^A in the new complex coordinates z^A will be given by

$$Z_k = \frac{1}{2}(\Lambda_k - i\Lambda_{\dot{k}}) = \frac{1}{2}\lambda_k, \qquad Z_{\dot{k}} = \frac{1}{2}(\Lambda_k + i\Lambda_{\dot{k}}) = \frac{i}{2}\lambda_{\dot{k}};$$

$$Z^k = \Lambda^k + i\Lambda^{\dot{k}} = 2\lambda^k, \qquad Z^{\dot{k}} = \Lambda^k - i\Lambda^{\dot{k}} = -2i\lambda^{\dot{k}}.$$

The following relations will be needed in our discussion and may be verified without difficulty:

$$
(19.1)\qquad
\begin{aligned}
&X_k = 2\frac{\partial}{\partial z^k}, &\quad X_j\bar{x}^i &= \frac{\partial \bar{z}^i}{\partial z^j}, &\quad X_{\bar{k}} &= -2i\frac{\partial}{\partial z^{\bar{k}}},\\
&X_{\bar{k}}\bar{x}^{\dot{i}} = \frac{\partial \bar{z}^{\dot{i}}}{\partial z^{\bar{k}}}, &\quad \widehat{\left(\frac{\partial \bar{z}^{i}}{\partial z^k}\right)} &= \frac{\partial \bar{z}^{\dot{i}}}{\partial z^{\bar{k}}}, &\quad \widehat{\left(\frac{\partial^2 \bar{z}^{i}}{\partial z^j\,\partial z^k}\right)} &= \frac{\partial^2 \bar{z}^{\dot{i}}}{\partial z^{\dot{j}}\,\partial z^{\bar{k}}}.
\end{aligned}
$$

The symbol $\frown$ designates the complex conjugate. Under the transformation $\bar{z}^k = \bar{z}^k(z^i)$, $\bar{z}^{\bar{k}} = \bar{z}^{\bar{k}}(z^{\dot{i}})$, the z components Z_A and Z^A of Λ_A and Λ^A transform by

$$
\bar{Z}_k = Z_r\frac{\partial z^r}{\partial \bar{z}^k}, \qquad \bar{Z}_{\bar{k}} = Z_{\dot{r}}\frac{\partial z^{\dot{r}}}{\partial \bar{z}^{\bar{k}}}; \qquad \bar{Z}^k = Z^r\frac{\partial \bar{z}^k}{\partial z^r}, \qquad \bar{Z}^{\bar{k}} = Z^{\dot{r}}\frac{\partial \bar{z}^{\bar{k}}}{\partial z^{\dot{r}}}
$$

and by merely replacing Z_A with λ_A and Z^A with λ^A these same equations give the transformation laws of the factor components.

We now come to the selection of an affine connection. Let us first consider a general affine connection $L^A{}_{BC}(z^D)$. Because of the separated form of the transformations $z^A \to \bar{z}^A$ all components of affine connection will transform as complex tensors with the exception of the sets $L^i{}_{jk}$ and $L^{\dot{i}}{}_{j\bar{k}}$ which will obey

$$
\bar{L}^r{}_{jk}\frac{\partial z^i}{\partial \bar{z}^r} = L^i{}_{rs}\frac{\partial z^r}{\partial \bar{z}^j}\frac{\partial z^s}{\partial \bar{z}^k} + \frac{\partial^2 z^i}{\partial \bar{z}^j\,\partial \bar{z}^k}, \qquad \bar{L}^{\dot{r}}{}_{j\bar{k}}\frac{\partial z^{\dot{i}}}{\partial \bar{z}^{\dot{r}}} = L^{\dot{i}}{}_{\dot{r}\dot{s}}\frac{\partial z^{\dot{r}}}{\partial \bar{z}^j}\frac{\partial z^{\dot{s}}}{\partial \bar{z}^{\bar{k}}} + \frac{\partial^2 z^{\dot{i}}}{\partial \bar{z}^j\,\partial \bar{z}^{\bar{k}}}.
$$

From (19.1) it follows that $\widehat{(L^i{}_{jk})} - L^i{}_{j\bar{k}}$ is a tensor and therefore we make the invariant restrictions

$$
(\mathrm{R}_{19.2})\qquad L^i{}_{j\bar{k}} = \widehat{(L^i{}_{jk})}.
$$

In selecting factor components of affine connection $l^A{}_{BC}$ we retain the restrictions of section 7 leaving us with but two sets of components $l^i{}_{jk}$ and $l^{\dot{i}}{}_{j\bar{k}}$ transforming by

$$
(5.3)\qquad l^r{}_{jk}\bar{X}_r x^i = l^i{}_{rs}\bar{X}_j x^r\bar{X}_k x^s + \bar{X}_j\bar{X}_k x^i \qquad l^{\dot{r}}{}_{j\bar{k}}\bar{X}_{\dot{r}} x^{\dot{i}} = l^{\dot{i}}{}_{\dot{r}\dot{s}}\bar{X}_j x^{\dot{r}}\bar{X}_{\bar{k}} x^{\dot{s}} + \bar{X}_j\bar{X}_{\bar{k}} x^{\dot{i}}
$$

yet to be defined. From (19.1) it follows that $l^i{}_{jk} - 2L^i{}_{jk}$ and $l^{\dot{i}}{}_{j\bar{k}} + 2iL^{\dot{i}}{}_{j\bar{k}}$ are tensors and therefore we define the only non-vanishing factor components of connection by

$$
(\mathrm{R}_{19.3})\qquad l^i{}_{jk} = 2L^i{}_{jk} \qquad l^{\dot{i}}{}_{j\bar{k}} = -2iL^i{}_{j\bar{k}}.
$$

The formulas (7.3) for the covariant derivatives of factor vectors are now simply multiples of the covariant derivatives of the complex vectors Z_k, Z^k and their complex conjugates. The restriction $(\mathrm{R}_{19.2})$ identifies the complex conjugate of the covariant derivative of a vector Z with the covariant derivative of the complex conjugate vector $\hat{Z}$.

The complex decomposition of this section is valid for the full complex group,

for the criteria for admissible transformations $x \to \bar{x}$ are $| X_{;}\bar{x}^i | = \left| \frac{\partial \bar{z}^i}{\partial z^j} \right| \neq 0$ and $| X_{\hat{;}}\bar{x}^i | = \left| \widehat{\frac{\partial \bar{z}^i}{\partial z^j}} \right| \neq 0$.

PRINCETON UNIVERSITY

REFERENCES

(1) L. P. Eisenhart and M. S. Knebelman, "Invariant theory of homogeneous contact transformations", Annals of Math., 37 (1936), 747–765.

(2) J. A. Schouten, "Zur Differentialgeometrie der Gruppe der Berührungstransformationen. I. Doppelthomogenen Behandlung vor Berührungstransformationen," Proc. Kon. Akad. Wet., Amsterdam, vol. 40 (1937) p. 100.

(3) F. Engel and K. Faber, "Die Liesche Theorie der partiellen Differentialgleichungen erster Ordnung", B. G. Teubner, 1932, p. 38.

(4) L. P. Eisenhart, "Contact transformations", Annals of Math., 30 (1928), p. 235.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

REPRESENTATION OF ERGODIC FLOWS

BY WARREN AMBROSE

(Received December 28, 1940)

Introduction

The theory of measure preserving transformations and flows (a flow is a 1-parameter group of measure preserving transformations) originated in the study of classical dynamical systems. The fact that von Neumann's proof of the mean ergodic theorem and G. D. Birkhoff's proof of the ergodic theorem make no use of the strong regularity conditions fulfilled by classical dynamical systems together with the fact that these theorems have applications in other fields where such strong regularity conditions are not fulfilled (notably in probability theory) has however led to a study of measure preserving transformations in terms of purely measure-theoretic formulations.[1] The present paper is concerned with the theory of flows from this measure-theoretic standpoint.

The principle result of this paper (Theorem 2) is a theorem which asserts that every (measurable) ergodic flow (an ergodic flow is one in which every set of positive measure sweeps out the whole space) is isomorphic (with respect to all measure properties) to a certain kind of flow that we call a flow "built under a function."[2] This representation theorem has various consequences of which the most important probably is the following: In studying the flows of classical dynamics one is constantly picking out cross sections, drawing tubes about the trajectories and, generally speaking, considering separately what happens along and what happens perpendicular to the direction of flow. This isomorphism theorem makes possible a similar analysis,—for an ergodic flow,—[3] under the most general measure theoretic formulations. It makes possible such an analysis because for a flow built under a function there exist certain sets with the properties of a cross section and the measure on the space is the direct product measure of a certain measure on the cross section with Lebesgue measure along the trajectories; these are the essential features necessary for such an analysis. As in the classical cases various properties of the flow become intimately tied up with the properties of a related transformation on this cross section. We re-

[1] See for example III, IV and X in the bibliography at the end of this paper.

[2] Flows of this kind were first introduced in a special case, by von Neumann in X, p. 636.

[3] Recently Prof. S. Kakutani has succeeded in extending this theorem to non-ergodic flows, using Theorem 1 of this paper. His method is similar to that which we use for the ergodic case. This and some other similar theorems will appear in a joint paper by Prof. Kakutani and the present author under the title, "Structure and Continuity of Measurable Flows." I take this opportunity to express my indebtedness to Prof. Kakutani for very many very helpful discussions of all the matters with which this paper is concerned.

mark that this representation theorem also enables us to show (except for one trivial and highly uninteresting kind of situation) that the trajectory of any point is always a measurable set (this we prove even for non-ergodic flows).

In section 1 we give our definitions and in section 2 we prove our general representation theorem. In section 3 we discuss the group of unitary operators always (since Koopman, V) associated with a flow and establish a spectral condition that a flow have a special kind of representation as a flow built under a function. We conclude by extending a theorem of von Neumann about differentiable ergodic flows to measurable ergodic flows; this extension is possible just because our isomorphism theorem enables us to obtain in the general case the tubes and cross sections for whose existence von Neumann assumed differentiability.

1. Definitions

Definition 1. A *measure space* is a space, Ω, on which a completed[4] (countably additive) measure, $m(M)$, is defined, for which $0 < m(\Omega) < \infty$ and such that there exists at least one measurable set, M, for which $0 < m(M) < m(\Omega)$.

Definition 2. A *measure preserving transformation*, T, is a 1:1 point transformation of a measure space Ω_1 onto[5] a measure space Ω_2 with the property that $M_2 = TM_1$ is measurable (in Ω_2) if and only if M_1 is measurable (in Ω_1) and for which $m_1(M_1) = m_2(M_2)$ for every measurable set M_1.[6]

Definition 3. A *flow* is a 1-parameter family, T_t, $(-\infty < t < \infty)$ of measure preserving transformations of a measure space onto itself, which has the group property: $T_tT_s = T_{t+s}$[7] for all t and s.

Definition 4. Let T_t be a flow on Ω. The set $M \subset \Omega$ is *invariant* under T_t if whenever P is in M then T_tP is in M for all t.

Definition 5. The flow T_t is *ergodic* if there are no measurable sets invariant under T_t except possibly sets of measure 0 and complements of sets of measure 0.

Definition 6. Let T_t be a flow on Ω' and let S_t be a flow on Ω''. T_t and S_t are *isomorphic* if it is possible to split Ω' into two disjoint measurable sets whose sum is Ω', Ω_1' and Ω_2', and to split Ω'' into two disjoint measurable sets whose sum is Ω'', Ω_1'' and Ω_2'', in such a way that:

1) Each of Ω_1' and Ω_2' is invariant under T_t and each of Ω_1'' and Ω_2'' is invariant under S_t.

2) Both Ω_2' and Ω_2'' have measure 0.[8]

[4] The measure $m(M)$ is completed if every subset of each set of measure 0 is measurable.

[5] We use the term "onto" with the customary meaning, that every point of Ω_{22} is the image of some point in Ω_1.

[6] Here we use m_1 for measure on Ω_1 and m_2 for measure on Ω_2. Obviously T is a measure preserving transformation of Ω_1 onto Ω_2 if and only if T^{-1} is a measure preserving transformation of Ω_2 onto Ω_1. Usually Ω_1 and Ω_2 will be the same.

[7] If T and S are transformations then by "$T = S$" we mean that $TP = SP$ for *every* point P.

[8] We allow the possibility that one or both of Ω_2' and Ω_2'' is empty.

3) There exists a measure preserving transformation, R, of Ω_1' onto Ω_1'' such that $T_tP = R^{-1}S_tRP$ for all P in Ω_1' (and all t).

Throughout this paper L will denote the real line taken with Lebesgue measure. If Ω is a measure space then $\Omega \times L$ will denote the product space of Ω with L, taken with the completed direct product measure defined in terms of Lebesgue measure on L and the given measure on Ω.

DEFINITION 7. Let T_t be a flow on Ω. T_t is *measurable* if the function T_tP is a measurable (P, t)-function, i.e. if for every measurable set, M, in Ω the (P, t)-set for which T_tP is in M is measurable in $\Omega \times L$.[9]

Equivalent to this definition is the assertion that for every complex-valued measurable function $f(P)$ defined on Ω the function $f(T_tP)$ is a measurable (P, t)-function.

DEFINITION 8. Let Ω be a measure space and T a measure preserving transformation of Ω onto itself. Consider the product space $\Omega \times L$ of Ω and the real line L (where measure on this product space is defined multiplicatively in terms of Lebesgue measure on L and the given measure on Ω). Let $f(P)$ be a real valued integrable function defined on Ω with $f(P) > c > 0$ for all P in Ω, and let $\bar{\Omega}$ be the portion of $\Omega \times L$ under the graph of $f(P)$, i.e. let $\bar{\Omega}$ be the set of points $\bar{P} = (P, x)$ for which $0 \leqq x < f(P)$. Then $\bar{\Omega}$ is a measure space. Define the flow T_t on $\bar{\Omega}$ by

$$(1.1)\quad \begin{cases} T_t(P, x) = (T^nP, t + x - f(P) - \cdots - f(T^{n-1}P)) \\ \qquad \text{for } n > 0 \text{ and } f(P) + \cdots + f(T^{n-1}P) - x \\ \qquad\qquad\qquad \leqq t < f(P) + \cdots + f(T^nP) - x, \\ T_t(P, x) = (P, t + x) \text{ for } -x \leqq t < f(P) - x, \\ T_t(P, x) = (T^{-n}P, t + x + f(T^{-1}P) + \cdots + f(T^{-n}P)) \\ \qquad \text{for } n > 0 \text{ and } -f(T^{-1}P) - \cdots - f(T^{-n}P) - x \\ \qquad\qquad\qquad \leqq t < -f(T^{-1}P) - \cdots - f(T^{-n+1}P) - x. \end{cases}$$

We call T_t the flow *built on the measure preserving transformation T under the function $f(P)$*.[10]

In this definition we have demanded that $f(P)$ be integrable only to make $\bar{\Omega}$ have finite measure; if we were considering flows on spaces of infinite measure then we would only have demanded that $f(P)$ be measurable and > 0. We might have required that $f(P)$ be positive everywhere rather than uniformly positive; this distinction is unessential but with $f(P)$ uniformly positive some of our proofs are simpler.

---

[9] In case Ω is a topological space and the measure is "properly" related to the topology it is usual to define a measurable flow as one with the property that for every open set 0 in Ω the (P, t)-set for which $T_tP \in 0$ is measurable in $\Omega \times L$. This apparently less restrictive definition is however equivalent to ours. For a proof see IV, pp. 9, 10.

[10] We shall not prove that the T_t thus defined actually are measure preserving transformations. A proof that they carry measurable sets into measurable sets is contained in the proof of Theorem 1. An application of Fubini's theorem then shows that they are measure preserving.

In considering flows built under a function it will be essential to consider the functions $F(\bar{P})$ and $G(\bar{P})$ defined by

$$F(\bar{P}) = F(P, x) = f(P),\text{[11]} \tag{1.2}$$

$$G(\bar{P}) = G(P, x) = x. \tag{1.3}$$

Intuitively, $G(\bar{P})$ is the length of time since a particle now at $\bar{P}$ was last in Ω, while $F(\bar{P})$ is the length of time since it was last in Ω plus the length of time before it will again be in Ω.

The sets defined by $G(\bar{P}) =$ constant give us cross sections for such a flow.

If M is any set in Ω we call the $\bar{P}$-set defined by

$$M^* = \underset{(P,x)}{[}P \in M]$$

a *tube* and we say that M^* is based on M. If M^* is a tube with $F(\bar{P}) \geqq d > 0$ for every $\bar{P}$ in M^* and if $0 \leqq a < b \leqq d$ we call the set $M^*(a, b)$[12] defined by

$$M^*(a, b) = M^* \cdot \underset{\bar{P}}{[}a \leqq G(\bar{P}) < b]$$

a *box*, and the set $M^*(b)$ defined by

$$M^*(b) = M^* \underset{\bar{P}}{[}G(\bar{P}) \geqq b]$$

a *leftover* of the tube. We say that M^* is the tube *through* the box $M^*(a, b)$.[13]

Let M^* be a tube and let $\bar{M}_k$ and $\bar{M}_{k,j}$ be defined, for each positive integer n, by

$$\begin{aligned} \bar{M}_k &= \underset{\bar{P}}{[}k2^{-n} \leqq F(\bar{P}) < (k+1)2^{-n}] \cdot M^*, \qquad k = 0, 1, 2, \cdots \\ \bar{M}_{k,j} &= \underset{\bar{P}}{[}j2^{-n} \leqq G(\bar{P}) < (j+1)2^{-n}] \cdot \bar{M}_k \quad j = 0, 1, \cdots, k-1. \end{aligned} \tag{1.4}$$

It is readily verified that

$$\sum_{k=0}^{\infty} \sum_{j=0}^{k-1} \bar{M}_{k,j} \to \sum_{k=0}^{\infty} \bar{M}_k = M^* \qquad (n \to \infty). \tag{1.5}$$

This fact will frequently be useful in what follows because it shows that every tube is a limit of sums of boxes.

2. The first representation theorem

In our definition of a flow built under a function we specified that the measure on $\bar{\Omega}$ be the direct product measure of a measure on Ω with Lebesgue measure on the x-axis. Suppose, however, that we have a flow defined by (1.1) on a region $\bar{\Omega}$ without assuming that the measure on $\bar{\Omega}$ is such a direct product measure, and suppose that we do not even have a measure defined on Ω (so that,

---

[11] We shall use the notations $\bar{P}$ and (P, x) interchangeably for points in $\bar{\Omega}$.

[12] We shall adhere to this notation, M^* for a tube, and $M^*(a, b)$ for a box through M^*, throughout this paper.

[13] Obviously a box uniquely defines the tube through itself.

in particular, we are not assuming that $f(P)$ is a measurable function or that T is a measure preserving transformation). The following theorem shows that if the flow satisfies certain measurability conditions then it is possible to put a measure on Ω for which T will be a measure preserving transformation and $f(P)$ will be a measurable function and such that the given measure on $\bar{\Omega}$ is the direct product of this measure on Ω with Lebesgue measure on the x-axis.

THEOREM 1. *Let T_t be a flow defined by* (1.1) *on a region $\bar{\Omega}$, with a measure $\bar{m}(\bar{M})$ on $\bar{\Omega}$. If T_t is a measurable flow and if the functions $F(\bar{P})$ and $G(\bar{P})$ are both $\bar{m}$-measurable, then there exists a measure $m(M)$ on Ω for which $f(P)$ is a measurable function and T is a measure preserving transformation and such that $\bar{m}(\bar{M})$ is the completed direct product measure*[14] *of $m(M)$ on Ω with Lebesgue measure on the x-axis.*

PROOF: We first define a Borel field[15] $\mathfrak{M}$ of sets in Ω and put a measure, $m(M)$, on it. We define $\mathfrak{M}$ to be the collection of those sets, M, in Ω with the property that M^*, the tube based on M, is $\bar{m}$-measurable. Obviously, $\mathfrak{M}$ is a Borel field. It is trivial from the fact that $F(\bar{P})$ is $\bar{m}$-measurable that $f(P)$ is measurable with respect to $\mathfrak{M}$. We define $m(M)$, for $M \in \mathfrak{M}$ by

$$m(M) = (1/c)\bar{m}(M^*(0, c)).^{16}$$

We want now to show that $m(M)$ is a completed measure and that, with this measure on Ω, T is a measure preserving transformation.

Because $F(\bar{P})$ and $G(\bar{P})$ are $\bar{m}$-measurable we know that if a tube is $\bar{m}$-measurable then so are its boxes and leftovers. Now we show that if a box is $\bar{m}$-measurable so is the tube through it. If a box is $\bar{m}$-measurable then any sub-box which has the same tube through it is $\bar{m}$-measurable because any such sub-box is obviously the intersection of the original box with a transform (by some T_t) of the original box. Also, if a box $M^*(a, b)$ is $\bar{m}$-measurable then the box $M^*(0, b)$ is $\bar{m}$-measurable since the latter is obviously the sum of a finite number of transforms of the former. To show that the tube through the box $M^*(a, b)$ is $\bar{m}$-measurable we see from (1.4) and (1.5) that it is sufficient to show that (for large n) each $\bar{M}_{k,j}$ is $\bar{m}$-measurable; the $\bar{m}$-measurability of the $\bar{M}_{k,j}$ (for large n) follows from the fact that

$$\bar{M}_{k,j} = T_{j2^{-n}}(M^*(0, 2^{-n}) \cdot \underset{\bar{P}}{[}k2^{-n} \leqq F(\bar{P}) < (k+1)2^{-n}]).$$

[14] The region $\bar{\Omega}$ is not a product space unless $f(P)$ = constant so it cannot be precisely correct to say that $\bar{m}(\bar{M})$ is a direct product measure. Precisely expressed what we mean is that if we define $m'(M')$ on $\Omega \times L$ as the completed direct product measure of $m(M)$ on Ω and Lebesgue measure on L then the following is true: if $\bar{M} \subset \bar{\Omega}$ then $\bar{M}$ is $\bar{M}$-measurable if and only if it is m'-measurable and if it is $\bar{M}$-measurable then $\bar{m}(\bar{M}) = m'(\bar{M})$.

[15] By a Borel field we mean a collection of sets closed under the operations of complementation and countable addition.

[16] The notation $M^*(a, b)$ was introduced above. The number c here and in the remainder of this proof is some fixed positive number with the property that $f(P) > c$ for all P in Ω. It is easily seen (and in fact is a direct consequence of a later part of this proof) that $m(M)$ is independent of the particular choice of c.

It follows readily from what we have just shown that $m(M)$ is a completed measure. We shall now show that T preserves m-measure. Let $M \epsilon \mathfrak{M}$ and $N = TM$; we shall show that the $\bar{m}$-measurability of $M^*(0, c)$ implies the $\bar{m}$-measurability of $N^*(0, c)$ and that these two sets have the same $\bar{m}$-measure. This will prove that if $M \epsilon \mathfrak{M}$ then $TM \epsilon \mathfrak{M}$ and $m(M) = m(TM)$. A similar proof shows that T^{-1} also has these properties so this will complete the proof that T is m-measure preserving. We define, for each positive integer n, the sets

$$\bar{M}^k = M^* \cdot \underset{\bar{P}}{[}k2^{-n} \leqq F(\bar{P}) < (k+1)2^{-n}] \underset{\bar{P}}{[}k2^{-n} - c \leqq G(\bar{P}) < k2^{-n}].$$

It is readily verified that

$$\left.\begin{aligned} T_c \sum_{k=0}^{\infty} \bar{M}^k &\to N^*(0, c) \\ \sum_{k=0}^{\infty} T_{c-k/2^n} \bar{M}^k &\to M^*(0, c)^{17} \end{aligned}\right\} \qquad (n \to \infty).$$

Because M^* is $\bar{m}$-measurable it follows from this that $N^*(0, c)$ and hence N^* is $\bar{m}$-measurable. Because in each of these sums the summands are disjunct it follows that $\bar{m}(M^*(0, c)) = \bar{m}(N^*(0, c))$ and hence that T is m-measure preserving.

Now we consider the space $\Omega \times L$ and in this space the Borel field $\mathfrak{B}_*$ determined by all sets of the form $M \times (a, b)$ where $M \epsilon \mathfrak{M}$.[18] Defining measure multiplicatively in $\Omega \times L$[19] we have a measure, $m'(B_*)$, defined for all $B_* \epsilon \mathfrak{B}_*$. Completing $\mathfrak{B}_*$ with respect to this measure we obtain a Borel field $\mathfrak{M}_*$. Now we note that $\bar{\Omega} \epsilon \mathfrak{B}_*$; this follows from (1.5) applied to the tube $M^* = \bar{\Omega}$. We define $\mathfrak{B}'$ to be the Borel field of sets of the form $B_*\bar{\Omega}$ where $B_* \epsilon \mathfrak{B}_*$, and $\mathfrak{M}'$ to be the Borel field of sets of the form $M_*\bar{\Omega}$ where $M_* \epsilon \mathfrak{M}_*$.[20] Let $\bar{\mathfrak{M}}$ denote the Borel field of $\bar{m}$-measurable sets. Then what we want to show is that $\mathfrak{M}' = \bar{\mathfrak{M}}$ and that for sets $\bar{M} \epsilon \bar{\mathfrak{M}}$ we have $m'(\bar{M}) = \bar{m}(\bar{M})$.

We now define $\mathfrak{F}'$ to be the field[21] consisting of those sets which are finite sums of boxes and leftovers of $\bar{m}$-measurable tubes. Now the Borel field determined by $\mathfrak{F}'$ is $\mathfrak{B}'$.[22] To see this we note first that $\mathfrak{B}'$ is the Borel field deter-

---

[17] The sets $\bar{M}_k$ and $\bar{M}_{k,j}$ are defined in terms of a given tube which we here take to be the tube through $M^*(a, b)$.

[18] By the interval (a, b) we shall mean the interval taken with its left end point a but without its right end point b. The Borel field determined by a collection of sets is the smallest Borel field containing that collection.

[19] For a discussion of product measures in abstract spaces see VI.

[20] An equivalent definition of $\mathfrak{M}'$ would have been to say that it is the Borel field obtained by completing $\mathfrak{B}'$ with respect to m'-measure.

[21] By a field we mean a collection of sets closed under the operations of complementation and finite addition.

[22] For a discussion of fields, Borel fields, etc. see VI.

mined by all sets of the form $(M \times (a, b))\cdot\bar{\Omega}$ where $M \in \mathfrak{M}$.[23] Now each such set is easily seen to be the sum of a box of some $\bar{m}$-measurable tube and a leftover of some $\bar{m}$-measurable tube; on the other hand, every $\bar{m}$-measurable box, every $\bar{m}$-measurable tube and every leftover of an $\bar{m}$-measurable tube is in $\mathfrak{B}'$, so $\mathfrak{F}' \subset \mathfrak{B}'$. This proves that $\mathfrak{B}'$ is the Borel field determined by $\mathfrak{F}'$.

Both $F(\bar{P})$ and $G(\bar{P})$ are measurable with respect to $\mathfrak{B}'$; this is true for $F(\bar{P})$ because the $\bar{P}$-set for which $F(\bar{P}) > b$ is, for any b, an $\bar{m}$-measurable tube and such a tube we already know to be in $\mathfrak{B}'$. This is true for $G(\bar{P})$ because

$$\underset{\bar{P}}{[G(\bar{P}) < b]} = \underset{\bar{P}}{[F(\bar{P}) < b]} + \underset{\bar{P}}{[F(\bar{P}) \geqq b]}\,\underset{\bar{P}}{[G(\bar{P}) < b]}.$$

The first set on the right side of this equality is an $\bar{m}$-measurable tube and the second is an $\bar{m}$-measurable box, so $G(\bar{P})$ is measurable with respect to $\mathfrak{B}'$.

We shall now prove that $\mathfrak{M}' \subset \bar{\mathfrak{M}}$ and that if $\bar{M} \in \mathfrak{M}'$ then $m'(\bar{M}) = \bar{m}(\bar{M})$. We know that every $\bar{m}$-measurable box belongs to $\bar{\mathfrak{M}}$, and that every leftover of an $\bar{m}$-measurable tube belongs to $\bar{\mathfrak{M}}$; it follows that $\mathfrak{F}'$ and then $\mathfrak{B}'$ is included in $\bar{\mathfrak{M}}$. We shall show that if $\bar{M} \in \mathfrak{B}'$ then $m'(\bar{M}) = \bar{m}(\bar{M})$. This trivially implies that $\mathfrak{M}' \subset \bar{\mathfrak{M}}$ and that m'-measure and $\bar{m}$-measure agree on $\mathfrak{M}'$. Consider an $\bar{m}$-measurable box, $M^*(a, b)$; let n be a fixed positive integer and write

$$\begin{aligned} M^*(a, b) = M^*(a, a + [b - a]/n) &+ M^*(a + [b - a]/n, a + 2[b - a]/n) \\ &+ \cdots \\ &+ M^*(b - [b - a]/n, b). \end{aligned}$$

These sets are all disjunct and are transforms of one another (under members of the group T_t); so each has $\bar{m}$-measure equal to $(1/n)\bar{m}(M^*(a, b))$. It follows readily that for y real and $a < y < b$ we have

$$\bar{m}(M^*(a, y)) = [(y - a)/(b - a)]\bar{m}(M^*(a, b)). \tag{2.1}$$

From the definition of m'-measure we have

$$m'(M^*(0, c)) = c(1/c)\bar{m}(M^*(0, c)) = \bar{m}(M^*(0, c)). \tag{2.2}$$

Taken together (2.1) and (2.2) imply that $m'(\bar{M}) = \bar{m}(\bar{M})$ whenever $\bar{M}$ is an $\bar{m}$-measurable box. By (1.5) it follows that this is true whenever $\bar{M}$ is an $\bar{m}$-measurable tube and hence for all $\bar{M}$ in $\mathfrak{F}'$. Hence $m'(\bar{M}) = \bar{m}(\bar{M})$ for all $\bar{M} \in \mathfrak{B}'$.[24] Hence $\mathfrak{M}' \subset \bar{\mathfrak{M}}$ and for sets in $\mathfrak{M}'$ the measures m' and $\bar{m}$ agree.

[23] This is a consequence of the following trivial lemma: If $\Omega_1 \supset \Omega_2$, $\{F_\alpha\}$ is a collection of sets in Ω_1, $\mathfrak{B}_1$ is the Borel field determined by the F_α, and $\Omega_2 \in \mathfrak{B}_1$ then the Borel field of sets of the form $M_1\Omega_2$ where $M_1 \in \mathfrak{B}_1$ is the same as the Borel field determined by sets of the form $F_\alpha\Omega_1$.

[24] The fact that if two (countably additive) measures agree on a field then they agree on the Borel field it determines is best proved by use of the theorem (VI, p. 85) that the normal family determined by a field is the same as the Borel field determined by the field.

Now we shall prove that each T_t takes sets of $\mathfrak{B}'$ into sets of $\mathfrak{B}'$ and sets of $\mathfrak{M}'$ into sets of $\mathfrak{M}'$. Because we know that $\mathfrak{M}' \subset \bar{\mathfrak{M}}$ and that the T_t are $\bar{m}$-measure preserving it is sufficient to prove the former, and because the T_t form a group it is sufficient to prove this for $0 \leqq t < c$. Because the $\bar{m}$-measurable boxes are a determining collection for $\mathfrak{B}'$ it is sufficient to show that any $\bar{m}$-measurable box goes into a set of $\mathfrak{B}'$. Let then $M^*(a, b)$ be an $\bar{m}$-measurable box and t a fixed real number, with $0 \leqq t < c$. Let M^* be the tube through $M^*(a, b)$, M the base of M^*, $N = TM$ and N^* the tube based on N. Then it is easily seen that

$$T_tM^*(a, b) = M^* \cdot \underset{\bar{P}}{[}a + t \leqq G(\bar{P}) < b + t]$$
$$+ \lim_{n \to \infty} \sum_{k=0}^{\infty} N^* \cdot \underset{\bar{P}}{[}k2^{-n} \leqq b + t - F(T_{-t}\bar{P}) < (k + 1)2^{-n}]$$
$$\cdot \underset{\bar{P}}{[}k2^{-n} + a - b \leqq G(\bar{P}) < k2^{-n}].$$

Now M^* and N^* are $\bar{m}$-measurable tubes and hence in $\mathfrak{B}'$; the set for which $a + t \leqq G(\bar{P}) < b + t$ is in $\mathfrak{B}'$ because $G(\bar{P})$ is measurable with respect to $\mathfrak{B}'$, and the sets under the summation are $\bar{m}$-measurable boxes and hence are in $\mathfrak{B}'$. Consequently $T_tM^*(a, b) \in \mathfrak{B}'$.[25]

It remains to show that $\bar{\mathfrak{M}} \subset \mathfrak{M}'$. Define the set $\bar{H}$ by

$$\bar{H} = \underset{\bar{P}}{[}G(\bar{P}) < c].$$

To show that $\bar{\mathfrak{M}} \subset \mathfrak{M}'$ it will be sufficient to show that if $\bar{M} \in \bar{\mathfrak{M}}$ and $\bar{M} \subset \bar{H}$ then $\bar{M} \in \mathfrak{M}'$. This is sufficient because any set in $\bar{\mathfrak{M}}$ is obviously the sum of a countable number of transforms of sets in $\bar{\mathfrak{M}}$ which are included in $\bar{H}$ and we know that each T_t takes sets of $\mathfrak{M}'$ into sets of $\mathfrak{M}'$.

Now we consider the product space $\bar{\Omega} \times L$ (L is the real line and $\mathfrak{L}$ will denote the Lebesgue measurable sets) and in this product space four Borel fields: $\mathfrak{M}' \times \mathfrak{L}$, $\mathfrak{M}' \otimes \mathfrak{L}$, $\bar{\mathfrak{M}} \times \mathfrak{L}$, $\bar{\mathfrak{M}} \otimes \mathfrak{L}$. These are defined as follows: $M' \times \mathfrak{L}$ is the Borel field determined by sets of the form $M \times E$, where $M \in \mathfrak{M}'$ and $E \in \mathfrak{L}$; $\mathfrak{M}' \otimes \mathfrak{L}$ is the completion of this Borel field with respect to the multiplicative measure defined on it. $\bar{\mathfrak{M}} \times \mathfrak{L}$ and $\bar{\mathfrak{M}} \otimes \mathfrak{L}$ are defined similarly in terms of $\bar{\mathfrak{M}}$ and $\mathfrak{L}$. From the fact that $\mathfrak{M}' \subset \bar{\mathfrak{M}}$ it follows that $\mathfrak{M}' \otimes \mathfrak{L} \subset \bar{\mathfrak{M}} \otimes \mathfrak{L}$ and that if $\tilde{M} \in \mathfrak{M}' \otimes \mathfrak{L}$ then the measure defined for it as a set of $\mathfrak{M}' \otimes \mathfrak{L}$ is the same as the measure defined for it as a set of $\bar{\mathfrak{M}} \otimes \mathfrak{L}$. We shall denote this measure in $\bar{\Omega} \times L$ by $\tilde{m}(\tilde{M})$.

In $\bar{\Omega} \times L$ we are considering points (P, x, t). If $\bar{M}$ is a (P, x)-set and $\bar{M}_1$

[25] What we are proving here is that a flow defined by (1.1), with measure taken as a product measure, takes measurable sets into measurable sets.

is a (P, t)-set and if $\bar{M}$ contains the same pairs (P, y) as $\bar{M}_1$ we shall write $\bar{M} = \bar{M}_1$. If $\tilde{M} \subset \bar{\Omega} \times L$ we shall use the following notations:

$$\tilde{M}_{(P,x)} = \underset{t}{[}(P, x, t) \in \tilde{M}]$$

$$\tilde{M}^t = \underset{(P,x)}{[}(P, x, t) \in \tilde{M}]$$

$$\tilde{M}_x = \underset{(P,t)}{[}(P, x, t) \in \tilde{M}].$$

Now we consider the transformation S of $\bar{\Omega} \times L$ onto itself defined by $S(P, x, t) = (P, x, t + x)$. This transformation takes sets of any one of our four Borel fields in $\bar{\Omega} \times L$ into sets of the same Borel field and preserves $\tilde{m}$-measure. We shall indicate the proof of this for $\bar{\mathfrak{M}} \times \mathfrak{L}$ and $\bar{\mathfrak{M}} \otimes \mathfrak{L}$, the proof for the other two being the same. First, if $\tilde{M} = \bar{M} \times (a, b)$ then obviously $S\tilde{M}$ is a limit of finite sums of such sets. Since S takes sums into sums and complements into complements it follows that S takes sets of $\bar{\mathfrak{M}} \times \mathfrak{L}$ into sets of $\bar{\mathfrak{M}} \times \mathfrak{L}$. The t-sets $\tilde{M}_{(P,x)}$ and $(S\tilde{M})_{(P,x)}$ are obviously congruent (for all (P, x)) and so have the same t-measure. Hence, by Fubini's theorem $\tilde{M}$ and $S\tilde{M}$ have the same $\tilde{m}$-measure. Because S preserves the measure of sets in $\bar{\mathfrak{M}} \times \mathfrak{L}$ it must take sets of $\bar{\mathfrak{M}} \otimes \mathfrak{L}$ into sets of $\bar{\mathfrak{M}} \otimes \mathfrak{L}$.

Now let $\bar{M} \in \bar{\mathfrak{M}}$, $\bar{M} \subset \bar{H}$, and consider the set $^*\tilde{M}$ defined by

$$^*\tilde{M} = (S \cdot \underset{(P,x,t)}{[}T_t(P, x) \in \bar{M}]) \left(\underset{(P,x,t)}{[} 0 \leqq t < c] \cdot \underset{(P,x,t)}{[} 0 \leqq x < c] \right).$$

Because T_t is a measurable flow and S is a measure preserving transformation in $\bar{\Omega} \times L$, $\tilde{M} \in \bar{\mathfrak{M}} \otimes \mathfrak{L}$. We shall call this $^*\tilde{M}$ the *associated set* of $\bar{M}$. It has the following readily established properties:

$$\text{(2.3)} \qquad \text{for } 0 \leqq x < c: \quad {}^*\tilde{M}_x = \bar{M}$$

$$\text{(2.4)} \qquad \text{for } 0 \leqq t < c: \quad {}^*\tilde{M}^t = \bar{H} \underset{(P,x)}{[} P \in \bar{M}_t].$$[26]

We now prove a sequence of assertions which will show that $\bar{\mathfrak{M}} \subset \mathfrak{M}'$. We number them for convenience.

1) If $\bar{M} \in \bar{\mathfrak{M}}$ and $^*\tilde{M}$ is its associated set then for almost all t, $^*\tilde{M}^t$ is in $\mathfrak{M}'$ and for almost all x, $\bar{M}_x$ is in $\mathfrak{M}$.

Let $\bar{M} \in \bar{\mathfrak{M}}$; then for almost all t, $^*\tilde{M}^t$ is $\bar{m}$-measurable (by Fubini's theorem) and for all t it is a box; hence $^*\tilde{M}^t \in \mathfrak{M}'$ for almost all t. If t is such that $^*\tilde{M}^t \in \mathfrak{M}'$ then it follows from (2.4) that $\bar{M}_t \in \mathfrak{M}$.[26] Hence $\bar{M}_x \in \mathfrak{M}$ for almost all x.

We now define $\tilde{\mathfrak{M}}$ to be the collection of all sets $\tilde{M}$ included in

$$\tilde{H} = \underset{(P,x,t)}{[} 0 \leqq x < c] \cdot \underset{(P,x,t)}{[} 0 \leqq t < c].$$

[26] $\bar{M}_t$ denotes the P-set: $\underset{P}{[}(P, t) \in \bar{M}]$.

and belonging to $\tilde{\mathfrak{M}} \otimes \mathfrak{L}$ which have the property that for almost all x, $\tilde{M}_x \in \mathfrak{M}'$.

2) If $\tilde{M} \subset \tilde{H}$ and $\tilde{M} \in \tilde{\mathfrak{M}} \times \mathfrak{L}$ then $\tilde{M} \in \tilde{\mathfrak{M}}$.

To prove this it is sufficient,—since $\tilde{\mathfrak{M}}$ is obviously a Borel field,[27]—to show that $\tilde{\mathfrak{M}}$ contains all sets of the form $\bar{M} \times (a, b)$ where $\bar{M} \in \bar{\mathfrak{M}}$, so consider such a set. We know from 1) that for almost all x, $\bar{M}_x$ belongs to $\mathfrak{M}$. Now $(\bar{M} \times (a, b))_x = \bar{\mathfrak{M}}_x \times (a, b)$ which is in $\mathfrak{M}'$ for almost all x.

3) If $\bar{M} \in \bar{\mathfrak{M}}$ (and $\bar{M} \subset \bar{H}$) then

$$\tilde{m}(\tilde{M}) = \int_0^c \bar{m}(\tilde{M}^t)\, dt = c \int_0^c m(\bar{M}_x)\, dx.$$

This is a consequence of Fubini's theorem and the fact that $\bar{m}(\tilde{M}^t) = cm(\bar{M}_t)$.

4) If $\bar{N} \in \bar{\mathfrak{M}}$ (and $\bar{N} \subset \bar{H}$) and has $\bar{m}$-measure 0 then $m(\bar{N}_x) = 0$ for almost all x in $(0, c)$.

If $\bar{N}$ has $\bar{m}$-measure 0 then

$$\underset{(P,x,t)}{[\;} T_t(P, x) \in \bar{N}]$$

has $\tilde{m}$-measure 0 (by Fubini's theorem and the fact that T_t is measurable). Hence the associated set, $^*\tilde{N}$, of $\bar{N}$ has $\tilde{m}$-measure 0. Then from

$$0 = \tilde{m}(^*\tilde{N}) = c \int_0^c m(\bar{N}_x)\, dx,$$

we see that $\bar{N}_x$ has m-measure 0 for almost all x in $(0, c)$.

5) If $\tilde{M} \in \tilde{\mathfrak{M}}$ (and $\tilde{M} \subset \tilde{H}$) and $\tilde{m}(\tilde{M}) = 0$ then $\bar{m}(\tilde{M}_x) = 0$ for almost all x in $(0, c)$.

By definition we know that $\tilde{M}_x \in \mathfrak{M}'$ except for an x-set (which we call θ_1) of Lebesgue measure 0. Also $\tilde{M}_{(P,x)}$ has t-measure 0 except for a (P, x)-set in $\bar{\mathfrak{M}}$ of $\bar{m}$-measure 0. Denote this exceptional (P, x)-set by $\bar{N}$. Then, by 4), except for an x-set (which we call θ_2) of Lebesgue measure 0 we have $m(\bar{N}_x) = 0$. We shall now show that if $x \notin \theta_1 + \theta_2$ then $\bar{m}(\tilde{M}_x) = 0$. Consider such an x. Then $\tilde{M}_x \in \mathfrak{M}'$. Then $\bar{N}_x$ has m-measure 0 and for P outside $\bar{N}_x$ we know that $m(\tilde{M}_{(P,x)}) = 0$. Hence, applying Fubini's theorem to the set $\tilde{M}_x$ (using the fact that it is in $\mathfrak{M}'$ and has t-measure 0 for almost all P) we see that $\bar{m}(\tilde{M}_x) = 0$.

6) If $\tilde{M} \subset \tilde{H}$ and $\tilde{M} \in \tilde{\mathfrak{M}} \otimes \mathfrak{L}$ then $\tilde{M} \in \tilde{\mathfrak{M}}$.

We already (by 2)) know this to be true when $\tilde{M} \in \tilde{\mathfrak{M}} \times \mathfrak{L}$, so it will be sufficient to show that if $\tilde{M} \in \tilde{\mathfrak{M}} \times \mathfrak{L}$ and $\tilde{m}(\tilde{M}) = 0$ and if $\tilde{N} \subset \tilde{M}$ then $\tilde{N} \in \tilde{\mathfrak{M}}$. This is true because we know by 2) and 5) that for such an $\tilde{M}$ we have $\tilde{M}_x \in \mathfrak{M}'$ and $\bar{m}(\tilde{M}_x) = 0$ for almost all x, which implies that $\tilde{N}_x \in \mathfrak{M}'$ for almost all x.

We can now conclude our proof. Let $\bar{M} \in \bar{\mathfrak{M}}$, $\bar{M} \subset \bar{H}$. Then the associated set $^*\tilde{M} \in \tilde{\mathfrak{M}} \otimes \mathfrak{L}$ and $^*\tilde{M} \subset \tilde{H}$; hence by 6) $\tilde{M} \in \tilde{\mathfrak{M}}$, so $^*\tilde{M}_x \in \mathfrak{M}'$ for almost all x in $(0, c)$. Since $^*\tilde{M}_x = \bar{M}$ for all x in $(0, c)$ this means that $\bar{M} \in \mathfrak{M}'$. Thus $\bar{\mathfrak{M}} \subset \mathfrak{M}'$ and hence $\bar{\mathfrak{M}} = \mathfrak{M}'$.

[27] I.e. $\tilde{\mathfrak{M}}$ is a Borel field when only sets within $\tilde{H}$ are considered and complements are taken with respect to $\tilde{H}$.

. THEOREM 2. *Every measurable ergodic flow is isomorphic to a flow built under a function.*

PROOF: Let S_t be the flow, Ω^* the space on which it occurs, and $m^*(M^*)$ the measure on Ω^*. We want to find three things,—a space Ω, a transformation T, and a function $f(P)$,—such that the flow T_t defined in terms of them by (1.1) is isomorphic to the original flow. Then we shall be able to apply Theorem 1 to conclude that T_t is built under a function, i.e. that the measure involved is a direct product measure. The Ω we choose will be a certain subset of Ω^* obtained by a judicious picking of sequences of points along the trajectories of S_t, the transformation T on Ω will be the one taking each point of Ω into the next point along the trajectory (under S_t) which lies in Ω, and the $f(P)$ will be the length of time (measured in terms of S_t) it takes to move (under S_t) from P to TP.

Let M^* be an m-measurable set with $0 < m^*(M^*) < m^*(\Omega^*)$ and let $\varphi(P^*)$ be its characteristic function. For P^* fixed and outside a certain invariant (under S_t) set N_1^*, of m^*-measure 0 $\varphi(S_tP^*)$ is a measurable t-function. By a theorem of Wiener[28] we know that

$$\frac{1}{\epsilon}\int_0^{\epsilon} \varphi(S_tP^*)\,dt \to \varphi(P^*) \qquad (\epsilon \to 0),$$

for almost all P^*. Hence (applying Egoroff's theorem) it is possible to choose a number, a, with $0 \leqq a < 1$, such that for $\Phi(P^*)$ defined by

$$\Phi(P^*) = \frac{1}{a}\int_0^{a} \varphi(S_tP^*)\,dt$$ [29]

the P^*-sets

$$M_1^* = \underset{P^*}{[}\Phi(P^*) < \tfrac{1}{4}], \qquad M_2^* = \underset{P^*}{[}\Phi(P^*) > \tfrac{3}{4}]$$

both have positive m^*-measure. Also, for $P^* \notin N_1^*$ the t-function $\Phi(S_tP^*)$ is continuous. In fact it is obvious that, for $P^* \notin N_1^*$,

$$(2.5) \qquad |\Phi(S_tP^*) - \Phi(S_sP^*)| \leqq (2/a)\,|t - s|.$$

Applying the ergodic theorem we know that for P^* fixed and outside a certain invariant set N_2^* of m^*-measure 0 the trajectory S_tP^* will have points in common with each of M_1^* and M_2^* for arbitrarily large and negatively arbitrarily large t. Now we write $\Omega_1^* = \Omega^* - [N_1^* + N_2^*]$ and $\Omega_2^* = N_1^* + N_2^*$; then Ω_1^* and Ω_2^* are m^*-measurable invariant sets and $m^*(\Omega_2^*) = 0$. We define Ω by

$$\Omega = \Omega_1^* \cdot \underset{P^*}{[}\Phi(P^*) = \tfrac{1}{2} \quad\text{and}\quad \Phi(S_tP^*) > \tfrac{1}{2} \quad \text{for all } t \text{ in } 0 < t \leqq a/8].$$

For the moment we shall not prove that Ω is an m^*-measurable set; this follows from the fact, proved below, that $G(\bar{P})$ is an $\bar{m}$-measurable function.

[28] Wiener, XI, Theorem III′, p. 2.

[29] $\Phi(P^*)$ is obviously m^*-measurable. Throughout this proof a is a fixed number.

Now we show that for any P^* in Ω_1^* the trajectory S_tP^* has points in common with Ω for arbitrarily large t; the same proof shows this to be true also for arbitrarily negatively large t. If t_0 is any positive number then there exist t_1 and t_2 such that $t_0 < t_1 < t_2$ and $S_{t_1}P^* \in M_1^*$, $S_{t_2}P^* \in M_2^*$. Because $\Phi(S_tP^*)$ is continuous in t there exists a t in (t_1, t_2) for which $\Phi(S_tP^*) = \frac{1}{2}$. We choose t' to be the greatest such t in (t_1, t_2). Then $S_{t'}P^* \in \Omega$ because $\Phi(S_{t'}P^*) = \frac{1}{2}$ while (2.5) implies

$$|t_2 - t'| \geqq (a/2)\,|\Phi(S_{t_2}P^*) - \Phi(S_{t'}P^*)| > a/8,$$

and hence that $\Phi(S_tS_{t'}P^*) > \frac{1}{2}$ for all t in $0 < t \leqq a/8$.

We define T and $f(P)$ as follows: if $P \in \Omega$ then there is a smallest positive number, t_1, for which $S_{t_1}P \in \Omega$.[30] Then $TP = S_{t_1}P$ and $f(P) = t_1$. We note that $f(P) > a/8$ for all P in Ω.

Now consider the region $\bar{\Omega}$ under the graph of $f(P)$[31], i.e. the set of points (P, x) for which $0 \leqq x < f(P)$, and let T_t be defined on $\bar{\Omega}$ by (1.1). We establish a 1:1correspondence, R, between Ω_1^* and $\bar{\Omega}$ as follows: if $(P, x) \in \bar{\Omega}$ then $P^* = S_xP$ is its corresponding point in Ω_1^*. This is obviously a 1:1 correspondence which carries S_t into T_t, i.e. if $P^* = R(P, x)$ then $S_tP^* = RT_t(P, x)$ for all t. In other words, $S_t\bar{P} = R^{-1}T_tR\bar{P}$ for all $\bar{P}$ in $\bar{\Omega}$. Now let $\bar{m}(\bar{M})$ be the measure on $\bar{\Omega}$ carried over from Ω_1^* by R.[32] If we knew $\bar{m}$-measure to be the direct product measure of a measure on Ω (for which T was a measure preserving transformation and $f(P)$ a measurable function) with Lebesgue measure on the x-axis then our theorem would be proved. To show this it is sufficient, by Theorem 1, to show that the functions $F(\bar{P})$ and $G(\bar{P})$ are $\bar{m}$-measurable.

To show that $G(\bar{P})$ is m-measurable let b be any number $\geqq 0$; then

$$R\underset{\bar{P}}{[}G(\bar{P}) \leqq b] = \sum_{0 \leqq t \leqq b} S_t\Omega,$$

so we must show the set on the right side to be m^*-measurable. This is shown by the following equalities (in these formulae some of the indices range over a whole interval; we adopt the following convention: when such a sum or a product is primed then the summation or product is to be taken over all rational numbers in the interval plus the end points of the interval):

$$\sum_{0 \leqq t \leqq b} S_t\Omega$$
$$= \sum_{0 \leqq t \leqq b} S_t\{[\Phi(P^*) = \tfrac{1}{2}][\Phi(S_sP^*) > \tfrac{1}{2} \quad \text{for } 0 < s \leqq a/8]\}$$

[30] That there exists a positive number t_1 with $S_{t_1}P \in \Omega$ follows from the last paragraph; there will be a smallest one (and it will be $> (a/8)$) because if $P \in \Omega$ then none of the points S_tP, for $0 < t \leqq (a/8)$, are in Ω.

[31] In this discussion we sometimes consider Ω as a subspace of Ω^* and sometimes as a space by itself.

[32] I.e. a set $\bar{M}$ is $\bar{m}$-measurable if and only if its correspondent in Ω_1 is m^*-measurable and its measure, $\bar{m}(\bar{M})$, is the m^*-measure of that correspondent.

$$= \sum_{0 \leqq t \leqq b} \{[\Phi(S_{-t}P^*) = \tfrac{1}{2}][\Phi(S_{s-t}P^*) > \tfrac{1}{2} \quad \text{for } 0 < s \leqq a/8]\}$$

$$= \prod_{n=1}^{\infty} {\sum_{0 \leqq t \leqq b}}' \{[|\Phi(S_{-t}P^*) - \tfrac{1}{2}| < 1/n][\Phi(S_{s-t}P^*) > \tfrac{1}{2} \quad \text{for } 0 < s \leqq a/8]\}$$

$$= \prod_{n=1}^{\infty} {\sum_{0 \leqq t \leqq b}}' \left\{[|\Phi(S_{-t}P^*) - \tfrac{1}{2}| < 1/n] \prod_{k=1}^{\infty} [\Phi(S_{s-t}P^*) > \tfrac{1}{2} \quad \text{for } 1/k \leqq s \leqq a/8]\right\}$$

$$= \prod_{n=1}^{\infty} {\sum_{0 \leqq t \leqq b}}' \left\{[|\Phi(S_{-t}P^*) - \tfrac{1}{2}| < 1/n] \prod_{k=1}^{\infty} \sum_{m=1}^{\infty} [\Phi(S_{s-t}P^*) \geqq \tfrac{1}{2} + 1/m \quad \text{for } 1/k \leqq s \leqq a/8]\right\}$$

$$= \prod_{n=1}^{\infty} {\sum_{0 \leqq t \leqq b}}' \left\{[|\Phi(S_{-t}P^*) - \tfrac{1}{2}| < 1/n] \prod_{k=1}^{\infty} \sum_{m=1}^{\infty} {\prod_{1/k \leqq s \leqq a/8}}' [\Phi(S_{s-t}P^*) \geqq \tfrac{1}{2} + 1/m]\right\},$$

which is obviously an m^*-measurable set.

To see that $F(\bar{P})$ is $\bar{m}$-measurable we note that

$$[F(\bar{P}) > b]_{\bar{P}} = {\sum_{0 \leqq t \leqq b}}' T_{-t}[G(\bar{P}) \geqq b]_{\bar{P}}.$$

We remarked in the introduction that Theorem 2 implies that, except for a trivial kind of exception, the trajectories in a measurable flow are always measurable sets. We shall now prove this. If S_t is any measurable flow on a space Ω^* then we can subdivide Ω^* into disjoint m^*-measurable invariant sets:

$$\Omega^* = \Omega_0^* + \Omega_1^* + \cdots + \Omega_n^* + \cdots,$$

(it may be that a finite or an infinite number of the Ω_n^* are empty; in fact for an ergodic flow all but one will be empty) in such a way that: 1) if $n \geqq 1$ then either Ω_n^* is empty or else $m^*(\Omega_n^*) > 0$ and S_t is ergodic on Ω_n^*, 2) either Ω_0^* is empty or else S_t is completely non-ergodic on Ω_0^*, i.e. if M^* is any measurable invariant set of positive measure then M^* contains an invariant set, N^*, for which $m^*(M^*) > m^*(N^*) > 0$. We see that every trajectory lying in Ω_0^* is measurable as follows: for each $\epsilon > 0$ it is clear that Ω_0^* can be divided into a finite number of disjoint measurable invariant sets (whose sum is Ω_0^*) each of measure $< \epsilon$. Each trajectory in Ω_0^* will lie in one of these sets. Hence each trajectory in Ω_0^* lies in some set of measure $< \epsilon$, and this for arbitrary positive ϵ. Consequently each trajectory in Ω_0^* lies in a set of measure 0 and therefore (since we always assume our measures to be completed) each trajectory in Ω_0^* is a measurable set of measure 0.

Since on each Ω_n^* (for $n \geqq 1$) which is non-empty S_t is ergodic (and since we took care of Ω_0^* in the last paragraph) the matter of the measurability of the trajectories will be cleared up completely if we can clear it up for ergodic flows. By Theorem 2 it will be sufficient (in considering ergodic flows) to consider flows built under a function. Let S_t be such a flow, on a region $\bar{\Omega}$, and let Ω, T, $f(P)$, $\bar{m}(\bar{M})$, $m(M)$ have their usual meanings. Clearly the ergodicity of S_t implies that

T is ergodic on Ω. We now distinguish three cases: 1) every point in Ω is m-measurable and of m-measure 0, 2) every point in Ω is m-measurable but some point has positive measure, 3) some point in Ω fails to be m-measurable. First we consider case 1) (the usual case). Let $S_t\bar{P}$ be any trajectory; this trajectory intersects Ω in a sequence of points $\bar{P}_n = (P_n, 0)$. Then this trajectory is the set

$$\sum_{n=-\infty}^{\infty} \underset{(P,x)}{[} P = P_n]$$

which is $\bar{m}$-measurable and of $\bar{m}$-measure 0, since our measure is a direct product measure. In case 2) the ergodicity of T implies that Ω consists of a finite number of points of equal positive m-measure plus perhaps a set of measure 0. Then S_t is obviously isomorphic to the well known periodic flow on the circle ($S_t x = x + t$ mod α, $\alpha > 0$). In this case all the trajectories are $\bar{m}$-measurable sets but a certain one of them has positive measure while all others have measure 0. In case 3) it is clear (since m-measure is completed) that the m-measurable sets will contain a "chunk," i.e. an m-measurable set, M, of positive m-measure and such that if N is m-measurable and contained in M then either $m(N) = m(M)$ or $m(N) = 0$. The ergodicity of T implies that M has a finite number of transforms and that every m-measurable set of positive m-measure differs by at most a set of m-measure 0 from some finite sum of transforms of M. Then our flow is essentially the same as in case 2) except that instead of all the measure being concentrated on the trajectory swept out by a single point all the measure is concentrated on the set swept out by M; clearly every $\bar{m}$-measurable set must be (to within a set of $\bar{m}$-measure 0) of the form:

$$\sum_{t \in E} S_t M,$$

where E is any Lebesgue measurable t-set on some finite interval $(0, \alpha)$. (Obviously in this case it is possible to set up a measure preserving *set*-correspondence between this and a flow of case 2) so in a certain sense this is the same as case 2)). In this case there will be at least one non-measurable trajectory inside the set swept out by M. This is the only case in which there can be any non-measurable trajectories and this case is obviously of not the slightest interest.

To summarize the results of this discussion we may say that if T_t is any measurable flow then the space on which it occurs can be split into a completely non-ergodic part, in which all trajectories have measure 0, plus a denumerable number of ergodic parts (where a finite or infinite number of these parts may be empty) and that on each ergodic part there are three possibilities: 1) each trajectory may be measurable and of measure 0, 2) each trajectory may be measurable but with one of the trajectories (necessarily a periodic one) having positive measure, in which case the flow is isomorphic to a rotation on the circle, 3) all the measure may be concentrated on the set swept out by a single set, M, of measure 0 in such a way that the only $\bar{m}$-measurable sets are sums of transforms of M; in this case non-measurable trajectories are possible, but it is possible

to establish a measure preserving *set* correspondence between such a flow and a rotation on the circle.

3. The group U_t and a second representation theorem

Let S_t be a flow on a space Ω^* and let L_2 be the space of (complex-valued) functions $g(P^*)$ of integrable square on Ω^*. We define the operator U_t (for all real t) by

(3.1) $$U_t g(P^*) = g(S_t P^*).[33]$$

Obviously the U_t are unitary operators and form a 1-parameter group: $U_t U_s = U_{t+s}$. It is known that if U_t is a continuous t-function (in the sense that, for any g in L_2, $\| U_t g - g \| \to 0$ as $t \to 0$) then U_t has a spectral resolution, i.e. there exists a unique spectral family E_λ for which

$$U_t = \int_{-\infty}^{\infty} e^{it\lambda}\, dE_\lambda,[34]$$

and it has been shown by Doob (III, Theorem 7) that if U_t is defined by (3.1), where S_t is a measurable flow, then U_t is a continuous t-function.[35] Hence U_t will have a spectral resolution in all cases we are considering.

We now define a special kind of flow built under a function and show that a measurable ergodic flow is isomorphic to a flow of this kind if and only if the group U_t has an eigenfunction of eigenvalue $\neq 0$. This result was proved,—except for the measure theoretic considerations,—by G. D. Birkhoff (II, p. 144) and our proof consists essentially in using his argument and then applying Theorem 2 to take care of these measure-theoretic considerations.

DEFINITION 9. An *eigenfunction* of the group U_t is a function $\psi(P^*)$ in L_2 which is different from 0 on some set of positive measure and such that for some real number λ,

$$U_t \psi = e^{i\lambda t} \psi$$

(for all real t). The number λ is called the *eigenvalue* of ψ.

DEFINITION 10. If a flow is built under a function $f(P)$, and $f(P)$ is a constant function[36] then we say the flow is *well built*. The number $f(P)$ is called the *height* of the flow.

[33] The introduction of this group of unitary operators and the idea of attempting to characterize the properties of a flow in terms of properties of the spectral resolution of this group of operators is due to B. O. Koopman, V.

[34] See VII and IX.

[35] If S_t is a measurable flow and U_t is defined by (3.1) then (by Fubini's theorem) $(U_t f, g)$ is a measurable t-function for each f and g in L_2 and Stone (VII) and von Neumann (IX) have shown that in case L_2 is separable this implies that U_t is a continuous t-function. In our case, however, where no separability conditions are assumed, it is necessary to use the deeper result of Doob to prove that U_t is a continuous t-function. We remark that for a measurable *ergodic* flow this continuity follows easily from Theorem 2.

[36] By constant we mean constant *everywhere*.

THEOREM 3. *A measurable ergodic flow is isomorphic to a well built flow if and only if its corresponding* U_t *group has an eigenfunction of eigenvalue* $\neq 0$. *The heights of the well built flows to which it is isomorphic are the numbers* $2\pi/|\lambda|$, *where* λ *is an eigenvalue of the* U_t.

PROOF: If T_t is a well built flow of height α and if λ is defined by $2\pi/\lambda = \alpha$ then obviously the function $\psi(P, x)$ defined by

$$\psi(P, x) = e^{i\lambda x}$$

is an eigenfunction of eigenvalue λ for the corresponding U_t. Since the eigenvalues form (for an ergodic flow) an additive group of real numbers[37] $-\lambda$ is also an eigenvalue. Hence if a measurable ergodic flow is isomorphic to a well built flow of height $2\pi/|\lambda|$ then both λ and $-\lambda$ are eigenvalues of the corresponding group U_t.

Now let S_t be a measurable ergodic flow on a space Ω^* and let the corresponding U_t have an eigenfunction ψ of eigenvalue λ, ($\lambda \neq 0$). Then $\bar{\psi}$ is an eigenfunction of eigenvalue $-\lambda$.[32] One of λ and $-\lambda$ is positive; we suppose λ is positive and from now on work with λ and ψ (if $-\lambda$ were the positive one we would work in the same way, but with $-\lambda$ and $\bar{\psi}$). Because ψ is an eigenfunction we know that for each t we have

$$\psi(S_t P^*) = e^{i\lambda t}\psi(P^*),$$

for almost all P^*. By changing $\psi(P^*)$ at most on a P^*-set of measure 0 we can make this equality hold for *all* P^* and t;[38] we suppose ψ to be so altered. Now $|\psi|$ is an invariant function (i.e. $|\psi(S_t P^*)| = |\psi(P^*)|$) and hence (since S_t is ergodic) $|\psi(P^*)| = \text{constant} = k > 0$ except for P^* in some m^*-measurable set, N^*, of m^*-measure 0.[39] Because $\psi(S_t P^*) = e^{i\lambda t}\psi(P^*)$ for all P^* and t we see that N^* is an invariant set. Now write $\Omega_2^* = N^*$ and $\Omega_1^* = \Omega^* - N^*$. Then define Ω by

$$\Omega = \Omega_1^* \cdot \underset{P^*}{[}\psi(P^*) = k]$$

(where k is the constant mentioned above). We establish a 1:1 correspondence between Ω_1^* and $\Omega \times (0, 2\pi/\lambda)$ as follows: if (P, x) is in $\Omega \times (0, 2\pi/\lambda)$ then make correspond to it the point P^* in Ω_1^* defined by $P^* = S_x P$. We define the transformation T on Ω by $TP = S_{2\pi/\lambda}P$. Then this correspondence obviously carries S_t on Ω_1^* into that flow on $\Omega \times (0, 2\pi/\lambda)$ which is built on the transformation T on Ω. As in the proof of Theorem 2 this correspondence gives rise to a measure, $\bar{m}(\bar{M})$, on Ω M $(0, 2\pi/\lambda)$ and the only thing remaining to be shown is that this measure is the desired sort of product measure. To prove this it is sufficient, by Theorem 1, to show that the functions $F(\bar{P})$ and $G(\bar{P})$ defined by (2.1) and (2.2) are $\bar{m}$-measurable. Now $G(\bar{P})$ is $\bar{m}$-measurable because the set in Ω_1^* corresponding to the $\bar{P}$-set for which $G(\bar{P}) < b$ is the P^*-set for which $\arg \psi(P^*) < b\lambda$. $F(\bar{P})$ is $\bar{m}$-measurable because it is a constant.

[37] See X, p. 625.

[38] See IV, Lemma 9.1, p. 27.

[39] This remark is due to Koopman, V, p. 318.

To conclude we point out that our first representation theorem can be used to generalize a theorem of von Neumann's about the spectral family E_λ associated with the group U_t. Let $\mathfrak{E}$ be the closed linear manifold spanned by the eigenfunctions of the group U_t. Then E_λ is still a spectral family when considered only on $L_2 - \mathfrak{E}$, the orthogonal complement of $\mathfrak{E}$[40]. The set of points of increase[41] of this spectral family when it is considered only on $L_2 - \mathfrak{E}$ is called the *band spectrum* of the original spectral family.

THEOREM 4.[42] *For a measurable ergodic flow the band spectrum is either the null set or the whole real line.*

PROOF: We shall not give a detailed proof of this theorem. We merely remark that von Neumann proved this assuming the flow to be differentiable in order to find,– in case the flow was not rotation on the circle,—cross sections whose translations over large time intervals were all disjoint. Using our representation theorem (and an unpublished but easily proved lemma of P. R. Halmos to the effect that on a space containing sets of arbitrarily small measure there exists, for every ergodic measure preserving transformation, T, and positive integer, n, a measurable set of positive measure whose first n transforms by T are all disjoint) we are able,—in case the flow is not rotation on the circle,—to obtain such cross sections without any differentiability assumptions. As soon as this is noticed von Neumann's proof can be carried through exactly as in the differentiable case.[43]

THE INSTITUTE FOR ADVANCED STUDY

BIBLIOGRAPHY

I. Birkhoff, G. D.: *Proof of the ergodic theorem*, P. N. A. S., v. 18, (1932), p. 650.
II. Birkhoff, G. D.: *Dynamical Systems*, A. M. A. Colloquium Pub., 1927.
III. Doob, J. L.: *One-parameter families of transformations*, Duke Jnl., v. 4, (1938), p. 752.
IV. Hopf, E.: *Ergodentheorie*, J. Springer, Berlin, 1937.
V. Koopman, B. O.: *Hamiltonian systems and linear transformations in Hilbert space*, P. N. A. S., v. 17, (1931), p. 315.
VI. Saks, S.: *Theory of the Integral*, G. E. Stechert, New York, 1938.
VII. Stone, M. H.: *On one parameter unitary groups in Hilbert space*, Ann. of Math., v. 33, (1932), p. 643.
VIII. Stone, M. H.: *Linear transformations in Hilbert space*, A. M. S. Colloquium Pub., 1932.
IX. von Neumann, J.: *Uber einen satz von Herrn M. H. Stone*, Ann. of Math., v. 33, (1932), p. 567.
X. von Neumann, J.: *Zur operatorenmethode in der klassischen mechanik*, Ann. of Math., v. 33, (1932), p. 587.
XI. Wiener, N.: *The ergodic theorem*, Duke Jnl., v. 5, (1939), p. 1.

[40] See VIII, Theorem 5.13, p. 189.

[41] See VIII, p. 184.

[42] See X, Theorem 5, p. 635.

[43] It has been pointed out to me by Dr. P. R. Halmos that this proof is not correct as it stands, but it has also been pointed out to me by Prof. S. Kakutani that it is possible to make it correct by some minor alterations in the definition of the function $\varphi(P)$.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

ON THE EXPONENTS OF DIFFERENTIAL IDEALS

BY ELLIS ROBERT KOLCHIN

(Received April 4, 1940)

INTRODUCTION

In the theory of polynomial ideals, in algebra, there are methods, stemming from the theorem of M. Noether, and associated with the names of E. Bertini, E. Lasker, F. S. Macauley, K. Hentzelt, H. Kapferer and P. Dubreil, for finding the exponent of an ideal, or at least a bound for the exponent.

When one seeks to create a notion of exponent for ideals of differential polynomials, one is forced, because of a situation revealed by H. W. Raudenbush,[1] to admit infinite exponents as well as finite ones. The investigation of such exponents, finite or infinite, to some extent for differential ideals of a general character, and to a deeper extent for differential ideals generated by a form in one unknown of the first order, is the object of the present paper.

If we may refer to §2 below for a definition of the *exponent of a differential ideal,* we shall proceed to enumerate our results.

Part II, which presents what is possibly the most interesting portion of our work, deals with differential ideals generated by a single form A, in one unknown, of the first order. The concept of *multiplicity* of a singular solution of A is introduced (§6) and it is shown (§7) that *if A has a singular solution of multiplicity exceeding unity, then the differential ideal generated by A has exponent infinity.* Forms A which have singular solutions, all of multiplicity unity, are discussed in §§8–10. Such singular solutions are divided into two classes, and, guided by a general theorem due to J. F. Ritt, we secure a decomposition of the differential ideal generated by A which puts these two classes into evidence (§8). In §10 it is proved, under an additional assumption (*regular type*) that *the exponent of the differential ideal generated by A is unity or two according as all the singular solutions are in the first class or at least one singular solution is in the second class.* The differential ideal generated by a form A which has *no* singular solutions, is shown, under a certain additional assumption, to have exponent *unity* (§11). Part II concludes with a discussion of a type of form which we call *hyperelliptic.* The "intermediate ideals" are found, and they are shown to fall into *chains* of a fixed length.

Part III contains a brief discussion of chains of differential ideals. A theorem is proved which gives a bound for the exponent of a so-called *principal* chain in terms of the length of the chain (§14).

Part I begins with a statement, in abstract form, of a decomposition theorem due to Ritt. After the definitions of *relative exponent* and other terms, there

[1] Bulletin of the American Mathematical Society, vol. 42 (1936), pp. 371–373.

is proved a theorem connecting the relative exponent of a differential ideal with those of the factor ideals in the decomposition just mentioned (Theorem 1). Also, a result is presented which relates the existence of a strong basis in a differential ideal to the existence of such a basis in the factor ideals (Theorem 2). The remainder of Part I deals with special questions whose treatment is important for the sequel.

In Part I, use is made of the theory of resolvents. This theory, as developed by Ritt, applies to a differential field whose elements are meromorphic functions. To permit the use of abstract fields, there is given, in an appendix, a discussion which supplements Ritt's proofs, making them valid in any differential field of characteristic zero.

Notation

Differentiation (of elements in a differential ring[2]) is indicated by subscripts; thus a_j means the j-th derivative of a. If a single letter is used with different subscripts to denote different ring elements, then differentiation is indicated by a second subscript; thus, the j-th derivative of a_i is a_{ij}.

Square brackets [ ], curled brackets { }, and parentheses ( ), when not used as symbols of aggregation, will mean, respectively, the differential ideal, the perfect differential ideal, and the (ordinary, algebraic) ideal generated by the set of elements they include. These various ideals are supposed to be formed in a fixed differential ring which underlies the discussion. When forming an (algebraic) ideal, the differential ring is considered as an (algebraic) ring.

Membership in a set is denoted in the usual way by the symbol ϵ. Set inclusion is indicated by $\subseteq$, proper inclusion by $\subset$. The symbol for the intersection of sets is $\cap$.

The notations

$$a \equiv b \qquad (m, n, \cdots),$$

$$a \equiv b \qquad [m, n, \cdots],$$

$$a \equiv b \qquad \{m, n, \cdots\}$$

will mean, respectively,

$$a - b \,\epsilon\, (m, n, \cdots), \qquad a - b \,\epsilon\, [m, n, \cdots], \qquad a - b \,\epsilon\, \{m, n, \cdots\}.$$

The product of a finite number of differential ideals is defined as their product when considered as (algebraic) ideals. This is readily seen to be a differential ideal. If $\sigma_1, \sigma_2, \cdots, \sigma_s$ are differential ideals, their product is denoted by $\sigma_1\sigma_2 \cdots \sigma_s$.

If σ is a *perfect* differential ideal in a differential ring $\mathfrak{D}$, and if $d \,\epsilon\, \mathfrak{D}$, then

[2] The definitions of differential ring, differential ideal, and other terms, are to be found in a paper by Raudenbush, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 361–368.

$\sigma : d$, the quotient of σ by d, is defined as in algebra to be the totality of elements $a \in \mathfrak{D}$ such that $ad \in \sigma$. $\sigma : d$ is itself a perfect differential ideal.

If $\mathfrak{F}$ is a differential field, then $\mathfrak{F}\{u, \cdots\}$ means the differential ring obtained by the differential ring adjunction of $u, \cdots$. The result of differential field adjunction will be denoted by $\mathfrak{F}\langle u, \cdots \rangle$.

As is customary, the word *form* will be used as an abbreviation for *differential polynomial.*

By a *solution* of a set of forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$, is meant a set of elements $\eta_1, \cdots, \eta_n$ of some differential extension field of $\mathfrak{F}$, which annul the forms of the set when substituted for $y_1, \cdots, y_n$, respectively.

Part I. Some General Theorems

1. The product representation

Consider a differential domain of integrity $\mathfrak{D}$ which contains the rational numbers. We suppose that every infinite system in $\mathfrak{D}$ has a basis. The set of forms in a fixed finite number of unknowns, with coefficients in an underlying differential field of characteristic zero, is such a differential domain of integrity.[3]

A set of differential ideals in $\mathfrak{D}$ will be called *separated* if $1 \equiv 0 \ (\sigma_1, \sigma_2)$ for each pair of distinct ideals σ_1, σ_2 in the set.

A differential ideal will be said to be *connected* if it is not the intersection of two separated differential ideals. Since the two ideals are required to be separated, "intersection" may be replaced by "product."[4]

A theorem due to Ritt may be formulated abstractly as follows:

Let σ be a differential ideal in $\mathfrak{D}$. Then σ is connected if and only if $\{\sigma\}$ is. If σ is not connected, it has a representation as the intersection (= product) of a separated finite set of connected differential ideals. This representation is unique. If $\sigma_1\sigma_2 \cdots \sigma_s$ is the representation of σ, then $\{\sigma_1\}\{\sigma_2\} \cdots \{\sigma_s\}$ is the representation for $\{\sigma\}$.[5]

We shall call the representation described in this theorem the *product representation* of σ. Each σ_i will be called a *factor* of the product representation.

2. Exponents

Lemma 1: *Let σ and τ be differential ideals in $\mathfrak{D}$ such that $\sigma \subseteq \tau \subseteq \{\sigma\}$. Then the product representation of σ has the same number of factors as that of τ, and, with a suitable assignment of subscripts, we may write for these representations*

$$(1) \qquad \begin{aligned} &\sigma = \sigma_1\sigma_2 \cdots \sigma_s, \qquad \tau = \tau_1\tau_2 \cdots \tau_s, \\ &\sigma_i \subseteq \tau_i \subseteq \{\sigma_i\}, \qquad\qquad i = 1, \cdots, s. \end{aligned}$$

Proof: The theorem of §1 shows that we may write for the product representations

[3] See the paper cited under footnote 2.

[4] See B. L. van der Waerden, *Moderne Algebra,* vol. 2, Berlin, 1931, p. 46.

[5] Ritt, Proceedings of the National Academy of Sciences, vol. 25 (1939), pp. 90–91.

$$\sigma = \sigma_1\sigma_2 \cdots \sigma_s, \qquad \tau = \tau_1\tau_2 \cdots \tau_s,$$
$$\{\sigma_i\} = \{\tau_i\}, \qquad i = 1, \cdots, s.$$

Here product is the same as intersection, and $\sigma \subseteq \tau \subseteq \tau_1$, so that

$$\sigma = (\sigma_1 \cap \tau_1) \cap \sigma_2 \cap \cdots \cap \sigma_s = (\sigma_1 \cap \tau_1)\sigma_2 \cdots \sigma_s.$$[6]

Since the product representation is unique, this implies that $\sigma_1 = \sigma_1 \cap \tau_1$, that is $\sigma_1 \subseteq \tau_1$. Similarly, every $\sigma_i \subseteq \tau_i$.

Consider two differential ideals σ and τ in $\mathfrak{D}$. If there is a positive integer q such that $\tau^q \subseteq \sigma$, let p be the least such q; if no such q exists, let $p = \infty$. We shall call p the *exponent of σ with respect to τ*, and shall denote it by $l_\tau\sigma$. The exponent of σ with respect to $\{\sigma\}$ will be called, simply, the *exponent of σ*.

In discussing $l_\tau\sigma$ we limit ourselves to cases for which $\sigma \subseteq \tau \subseteq \{\sigma\}$. If $\tau \not\subseteq \{\sigma\}$, then $l_\tau\sigma = \infty$; if $\sigma \not\subseteq \tau$, we may let $\tau' = (\sigma, \tau)$, and then $\sigma \subseteq \tau'$, $l_\tau\sigma = l_{\tau'}\sigma$.

THEOREM 1: *Let σ and τ be differential ideals in $\mathfrak{D}$ such that $\sigma \subseteq \tau \subseteq \{\sigma\}$. Let the product representations of σ and τ be given by* (1). *Then*

$$l_\tau\sigma = \max_{i=1,\cdots,s} l_{\tau_i}\sigma_i.$$

PROOF: Let $p = \max l_{\tau_i}\sigma_i$, $q = l_\tau\sigma$. If $p = \infty$, then $q \leqq p$. Suppose $p < \infty$. Then

$$\tau^p = (\tau_1 \cdots \tau_s)^p = \tau_1^p \cdots \tau_s^p \subseteq \sigma_1 \cdots \sigma_s = \sigma,$$

so that in either case $q \leqq p$. It remains to prove that $p \leqq q$.

If $q = \infty$, then $p \leqq q$. Suppose $q < \infty$. Then

$$\sigma_1^q \cdots \sigma_s^q = (\sigma_1 \cdots \sigma_s)^q = \sigma^q \subseteq \tau = \tau_1 \cdots \tau_s.$$

But $\sigma_1^q, \cdots, \sigma_s^q$ are obviously the factors of the product representation of $\sigma_1^q \cdots \sigma_s^q$. Hence, by Lemma 1, $\sigma_i^q \subseteq \tau_i$, $i = 1, \cdots, s$, that is, $p \leqq q$.

3. Strong bases

In this section we present a theorem which will not be used in the rest of the paper, but which may be of some interest on its own account.

Let σ be a differential ideal in $\mathfrak{D}$. If σ has a strong[7] basis $b_1, \cdots, b_h$, let $q = l_\sigma[b_1, \cdots, b_h]$; of all integers so obtained, let p be the minimum. If σ has no strong basis, let $p = \infty$. We shall call p the *basis index* of σ.

[6] It is easy to see that $\sigma_1 \cap \tau_1, \sigma_2, \cdots, \sigma_s$ are separated. For example, because $\sigma_1, \sigma_2, \cdots, \sigma_s$ are separated, there are elements $h_1 \in \sigma_1$ and $h_2 \in \sigma_2$ such that $1 = h_1 + h_2$. Choosing, as we may, p large enough so that $h_1^p \in \tau_1$, we have $h_1^p \in \sigma_1 \cap \tau_1$, $ph_1^{p-1}h_2 + \cdots + h_2^p \in \sigma_2$ and $1 = h_1^p + ph_1^{p-1}h_2 + \cdots + h_2^p$.

[7] A basis of σ is said to be *strong* if there is a single integer such that g^t is in the differential ideal generated by the basis, for every $g \in \sigma$. There exist differential ideals which do not have a strong basis. See Theorem 5 below; also, Kolchin, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 923–926.

THEOREM 2: *Let the differential ideal σ in $\mathfrak{D}$ have the product representation $\sigma = \sigma_1 \cdots \sigma_s$. Then the basis index of σ equals the maximum of the basis indices of the σ_i.*

PROOF: We know that there exist elements m_i, with derivative $m_{i1} \in \sigma$, such that $\sigma_i = (\sigma, m_i)$, $i = 1, \cdots, s$.[8] Let $b_1, \cdots, b_h$ be a strong basis of σ, with $l_\sigma[b_1, \cdots, b_h] = p$. Then

$$\sigma_i^p = (\sigma, m_i)^p \subseteq (\sigma^p, m_i) \subseteq [b_1, \cdots, b_h, m_i].$$

Thus, for each i, $b_1, \cdots, b_h, m_i$ is a strong basis for σ_i, and $l_{\sigma_i}[b_1, \cdots, b_h, m_i] \leqq p$. Hence the basis index of σ is not less than that of each σ_i. Now let $b_1^{(i)}, \cdots, b_{h_i}^{(i)}$ be a strong basis of σ_i, and let $q \geqq l_{\sigma_i}[b_1^{(i)}, \cdots, b_{h_i}^{(i)}]$, $i = 1, \cdots, s$. Since $\sigma_i = (\sigma, m_i)$, we may write, for each i and j,

$$b_j^{(i)} = t_j^{(i)} + c_j^{(i)} m_i, \qquad t_j^{(i)} \in \sigma.$$

Then

$$\begin{aligned} \sigma^q &= (\sigma_1 \cdots \sigma_s)^q = \sigma_1^q \cdots \sigma_s^q \\ &\subseteq [b_1^{(1)}, \cdots, b_{h_1}^{(1)}] \cdots [b_1^{(s)}, \cdots, b_{h_s}^{(s)}] \\ &\subseteq [t_1^{(1)}, t_2^{(1)}, \cdots, t_{h_s}^{(s)}, m_{11}, \cdots, m_{s1}, m_1 \cdots m_s]. \end{aligned}$$

Thus, σ has a strong basis whose associated exponent is less than or equal to q, so that the basis index of σ is not more than the greatest basis index of the σ_i.

4. Invariance of the exponent under differential field adjunction

In this section we specialize $\mathfrak{D}$ to a differential ring of forms.

Let $\mathfrak{F}$ be a differential field of characteristic zero. Let $\mathfrak{F}'$ be a differential extension field of $\mathfrak{F}$: $\mathfrak{F} \subseteq \mathfrak{F}'$. Throughout the present section y_i, u_i, w will denote unknowns.

If Φ is a set of forms with coefficients in $\mathfrak{F}$, we shall use (Φ), $[\Phi]$, $\{\Phi\}$ to denote the algebraic, the differential, and the perfect differential ideals, respectively, generated by Φ in the differential ring of forms with coefficients in $\mathfrak{F}$. If Φ is a set of forms with coefficients in $\mathfrak{F}'$, the algebraic, the differential, and the perfect differential ideals generated by Φ in the differential ring of forms with coefficients in $\mathfrak{F}'$ will be indicated, respectively, by $(\Phi)'$, $[\Phi]'$, and $\{\Phi\}'$. It is clear that if Σ is a differential ideal of forms with coefficients in $\mathfrak{F}$, then $(\Sigma)' = [\Sigma]'$.

THEOREM 3: *Let Σ and Λ be differential ideals in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then*

$$l_\Lambda \Sigma = l_{(\Lambda)'}(\Sigma)'.$$

PROOF: Let $p = l_\Lambda \Sigma$, $q = l_{(\Lambda)'}(\Sigma)'$. If $q = \infty$, then $p \leqq q$. Suppose $q < \infty$, and let $G \in \Lambda$. Then $G \in (\Lambda)'$, so that $G^q \in (\Sigma)'$. But $G^q \in \mathfrak{F}\{y_1, \cdots, y_n\}$. Hence $G^q \in \Sigma$,[9] so that $p \leqq q$.

[8] Ritt, loc. cit. under footnote 5.

[9] See proof in van der Waerden, loc. cit., p. 67.

If $p = \infty$, then $q \leqq p$. Suppose $p < \infty$, and let $G \in (\Lambda)'$. Then there is a relation

$$G = \sum_{\mu=1}^{m} H_\mu G_\mu, \qquad G_\mu \in \Lambda.$$

Hence

$$G^p = \sum_{i_1+\cdots+i_m=p} \frac{p!}{i_1! \cdots i_m!} H_1^{i_1} \cdots H_m^{i_m} G_1^{i_1} \cdots G_m^{i_m} \in (\Lambda^p)' \subseteq (\Sigma)',$$

so that $q \leqq p$. This completes the proof.

Theorem 3, by itself, yields no information as to how the exponent of a differential ideal is affected by a differential field adjunction, for there is no obvious reason why $\{\Sigma\}'$ should equal $(\{\Sigma\})'$. These two ideals, however, are identical. We devote the rest of this section to the proof of this fact. We use three lemmas.

LEMMA 2: *Let Ω be a prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, w, y_1, \cdots, y_p\}$, with basic set*

$$(2) \qquad A, A_1, \cdots, A_p$$

introducing in succession $w, y_1, \cdots, y_p$. Let A be algebraically irreducible over $\mathfrak{F}$, and let each A_i be of order zero and degree unity in y_i. Then (2) is a basic set for $\{\Omega\}'$.

PROOF: Let $B_1, \cdots, B_s$ be the irreducible factors of A over $\mathfrak{F}'$. Let S and I be the separant and initial, respectively, of A; let I_i be the initial (= separant) of A_i. Each B_i is of the same order (call it r) in w as A is, for otherwise A would be reducible over $\mathfrak{F}$. Let T_i be the separant, J_i the initial, of B_i. We have

$$I = J_1 \cdots J_s,$$

$$S = T_1 B_2 \cdots B_s + B_1 T_2 \cdots B_s + \cdots + B_1 B_2 \cdots T_s.$$

Now, $\Omega = \{A, A_1, \cdots, A_p\} : ISI_1 \cdots I_p$. Let

$$\Omega_i = \{B_i, A_1, \cdots, A_p\} : J_i T_i I_1 \cdots I_p, \qquad i = 1, \cdots, s.$$

Then each Ω_i is a prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, w, y_1, \cdots, y_p\}$.[10] Moreover,

[10] PROOF: Let $CD \in \Omega_i$. For appropriate k we may write $(I_1 \cdots I_p)^k C \equiv C'$, $(I_1 \cdots I_p)^k D \equiv D'$ $[A_1, \cdots, A_p]$ where C' and D' are forms free of $y_1, \cdots, y_p$. Of course $C'D' \in \Omega_i$. We prove one of C', D' is divisible by B_i. Suppose neither is. Then there are forms M and N, free of $y_1, \cdots, y_p$, such that $E = M(C'D'T_i J_i I_1 \cdots I_p) + NB_i$, where E is a nonzero form free of $y_1, \cdots, y_p$, of order less than r in w. Clearly $E \in \{B_i, A_1, \cdots, A_p\}$. It readily follows that $(T_i I_1 \cdots I_p)^l E \in (B_i)$ for some positive integer l. Since B_i is irreducible, this implies that $E \in (B_i)$. But the order of E in w is less than r. Hence $E = 0$. This contradiction shows that either C' or D' is divisible by B_i, say C' is. Then $C \in \Omega_i$. Thus, Ω_i is prime.

$$
\begin{aligned}
\{\Omega\}' &= \{\{A, A_1, \cdots, A_p\}:ISI_1 \cdots I_p\}' \\
&= \{A, A_1, \cdots, A_p\}':ISI_1 \cdots I_p \\
&= (\{B_1, A_1, \cdots, A_p\}' \cap \cdots \cap \{B_s, A_1, \cdots, A_p\}'):ISI_1 \cdots I_p \\
&= (\{B_1, A_1, \cdots, A_p\}':ISI_1 \cdots I_p) \cap \cdots \\
&\qquad\qquad \cap (\{B_s, A_1, \cdots, A_p\}':ISI_1 \cdots I_p) \\
&= (\{B_1, A_1, \cdots, A_p\}':J_1 \cdots J_s T_1 B_2 \cdots B_s I_1 \cdots I_p) \\
&\qquad\qquad \cap \cdots \cap (\{B_s, A_1, \cdots, A_p\}':J_1 \cdots J_s B_1 B_2 \cdots T_s I_1 \cdots I_p) \\
&= (\Omega_1:J_2 \cdots J_s B_2 \cdots B_s) \cap \cdots \cap (\Omega_s:J_1 \cdots J_{s-1} B_1 \cdots B_{s-1}) \\
&= \Omega_1 \cap \cdots \cap \Omega_s.
\end{aligned}
$$
[11]

$\{\Omega\}'$ does not contain a form of class $q + 1$ of order less than r in w, because Ω_i does not. Let C be of class $q + 1$ and have order r in w, and suppose that $C \in \{\Omega\}'$. $C \in \Omega_i$, and is therefore divisible by B_i, $i = 1, \cdots, s$. Hence C is divisible by A. From this it follows that (2) is a basic set for $\{\Omega\}'$.

LEMMA 3: *Let $\mathfrak{F}$ contain non-constant elements.*[12] *Let Σ be a prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, with arbitrary unknowns $u_1, \cdots, u_q$.*[13] *Then $(\Sigma)' = \{\Sigma\}'$.*

PROOF: It suffices to show that $\{\Sigma\}' \subseteq (\Sigma)'$. Let $A = 0$ be a resolvent for Σ, and let Ω be the associated prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, w, y_1, \cdots, y_p\}$, with basic set (2), where A is of order r in w, and each A_i is of order zero and degree unity in y_i.[14] We know that

$$\Sigma = \Omega \cap \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}.$$

Let $G \in \{\Sigma\}'$. We may write

$$G = G_0 + G_1\omega_1 + \cdots + G_m\omega_m$$

where each $G_\mu \in \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, and where $1, \omega_1, \cdots, \omega_m$ are elements of $\mathfrak{F}'$, linearly independent over $\mathfrak{F}$. We shall show that each $G_\mu \in \Sigma$.

Let S and I be the separant and initial of A; let I_i be the initial of A_i. Evidently there exist non-negative integers a, b, a_j such that

$$I^a S^b I_1^{a_1} \cdots I_p^{a_p} G_\mu \equiv R_\mu(\Omega), \qquad \mu = 0, 1, \cdots, m, \tag{3}$$

where each R_μ is reduced with respect to the basic set (2). Then

$$I^a S^b I_1^{a_1} \cdots I_p^{a_p} G \equiv R_0 + R_1\omega_1 + \cdots + R_m\omega_m (\Omega)'. \tag{4}$$

But, by Lemma 2, (2) is a basic set for $\{\Omega\}'$. Hence the second member of (4),

[11] The parentheses in these equations are symbols of aggregation.

[12] An element of $\mathfrak{F}$ is called *constant*, if its derivative vanishes.

[13] The u_i need not actually occur.

[14] See Appendix.

which is in $\{\Omega\}'$, yet reduced with respect to (2), must vanish. Since $1, \omega_1, \cdots, \omega_m$ are linearly independent over $\mathfrak{F}$, it follows that each R_μ vanishes. (3) then shows that each $G_\mu \in \Omega$. Since G_μ is also in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, we have $G_\mu \in \Sigma$, $\mu = 0, 1, \cdots, m$.

Hence $G \in (\Sigma)'$, so that $\{\Sigma\}' \subseteq (\Sigma)'$.

LEMMA 4: *Let $\mathfrak{F}$ consist purely of constants; let $\mathfrak{F}' = \mathfrak{F}\langle x\rangle$, where x is an element whose derivative is 1. Let Σ be a prime differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then $(\Sigma)' = \{\Sigma\}'$.*

PROOF: We must show that $\{\Sigma\}' \subseteq (\Sigma)'$. Let $G \in \{\Sigma\}'$. For a suitable non-zero element $\varphi \in \mathfrak{F}\langle x\rangle$, we may write

$$\varphi G = H_0 + H_1 x + \cdots + H_m x^m,$$

where the H_μ are in $\mathfrak{F}\{y_1, \cdots, y_n\}$. For a large integer r, $G^r \in (\Sigma)'$. Hence

$$H_0^r + rH_0^{r-1}H_1 x + \cdots + H_m^r x^{mr} \in (\Sigma)'.$$

Thus, we see that we may write

$$H_0^r + rH_0^{r-1}H_1 x + \cdots + H_m^r x^{mr} = d_1 S_1 + \cdots + d_k S_k,$$

where each $S_i \in \Sigma$, and each $d_i \in \mathfrak{F}\langle x\rangle$. Let $\omega_1, \cdots, \omega_t$ be elements of $\mathfrak{F}\langle x\rangle$ such that $1, x, \cdots, x^{mr}, \omega_1, \cdots, \omega_t$ are linearly independent over $\mathfrak{F}$, and such that each d_i is linearly dependent on $1, x, \cdots, x^{mr}, \omega_1, \cdots, \omega_t$. Then there is a relation

$$H_0^r + rH_0^{r-1}H_1 x + \cdots + H_m^r x^{mr} = T_0 + \cdots + T_{mr} x^{mr} + U_1\omega_1 + \cdots + U_t\omega_t,$$

where the T_i and U_i are in Σ. By the linear independence it follows that $H_0^r = T_0$, so that $H_0 \in \Sigma$. Hence

$$H_1 + H_2 x + \cdots + H_m x^{m-1} \in \{\Sigma\}'.$$

Continuing, we find every $H_\mu \in \Sigma$. Therefore $G \in (\Sigma)'$ and $\{\Sigma\}' \subseteq (\Sigma)'$.

We are now in a position to prove our theorem. $\mathfrak{F}$ is again any differential field of characteristic zero, and $\mathfrak{F}'$ a differential extension thereof.

THEOREM 4: *Let Σ be a perfect differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then $(\Sigma)' = \{\Sigma\}'$.*

PROOF: Let $\Sigma = \Sigma_1 \cap \cdots \cap \Sigma_s$ be the decomposition of Σ into essential prime differential ideals. It is easy to see that

$$\{\Sigma\}' = \{\Sigma_1\}' \cap \cdots \cap \{\Sigma_s\}'. \tag{5}$$

We shall prove that

$$(\Sigma)' = (\Sigma_1)' \cap \cdots \cap (\Sigma_s)'. \tag{6}$$

Indeed, since $\Sigma \subseteq \Sigma_i$, we have $(\Sigma)' \subseteq (\Sigma_i)'$, $i = 1, \cdots, s$, that is

$$(\Sigma)' \subseteq (\Sigma_1)' \cap \cdots \cap (\Sigma_s)'.$$

Now let $G \in (\Sigma_i)'$, $i = 1, \cdots, s$. We may write

$$G = G_0^{(i)} + G_1^{(i)}\omega_1 + \cdots + G_m^{(i)}\omega_m,$$

where each $G_i^{(j)} \epsilon \Sigma_j$, and $1, \omega_1, \cdots, \omega_m$ are elements of $\mathfrak{F}'$, linearly independent over $\mathfrak{F}$. The linear independence shows that $G_i^{(1)} = \cdots = G_i^{(s)}, i = 0, 1, \cdots, m$, so that each $G_i^{(j)} \epsilon \Sigma_1 \cap \cdots \cap \Sigma_s = \Sigma$, and $G \epsilon (\Sigma)'$. Hence

$$(\Sigma)' \supseteq (\Sigma_1)' \cap \cdots \cap (\Sigma_s)'.$$

This proves (6).

If $\mathfrak{F}$ contains non-constants, then, by (5), (6) and Lemma 3, $\{\Sigma\}' = (\Sigma)'$.

Suppose $\mathfrak{F}$ consists purely of constants. Let $\mathfrak{F}^* = \mathfrak{F}\langle x \rangle$, where x is an element whose derivative is 1, and let $\mathfrak{F}^\ddagger = \mathfrak{F}'\langle x \rangle$. By Lemma 4 and equations analogous to (5) and (6),

$$\begin{aligned}(\Sigma)^* &= (\Sigma_1)^* \cap \cdots \cap (\Sigma_s)^* \\ &= \{\Sigma_1\}^* \cap \cdots \cap \{\Sigma_s\}^* = \{\Sigma\}^*.\end{aligned}$$

Hence, by the part of the theorem already proved,

$$\begin{aligned}(\Sigma)^\ddagger &= ((\Sigma)^*)^\ddagger = (\{\Sigma\}^*)^\ddagger \\ &= \{\{\Sigma\}^*\}^\ddagger = \{\Sigma\}^\ddagger.\end{aligned}$$

Thus, $(\Sigma)^\ddagger$ is a perfect differential ideal.

Now, $(\Sigma)^\ddagger = ((\Sigma)')^\ddagger$. Suppose that $(\Sigma)' \neq \{\Sigma\}'$. Then there is an $F \epsilon \mathfrak{F}'\{y_1, \cdots, y_n\}$ such that $F \epsilon \{\Sigma\}'$, $F \not\epsilon (\Sigma)'$. Clearly $F \epsilon \{\Sigma\}^\ddagger$, that is, $F \epsilon (\Sigma)^\ddagger = ((\Sigma)')^\ddagger$. Since $F \epsilon \mathfrak{F}'\{y_1, \cdots, y_n\}$, this implies that $F \epsilon (\Sigma)'$. This contradiction shows that $(\Sigma)' = \{\Sigma\}'$, and completes the proof.

Corollary 1: *Let Σ be a differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then the exponent of Σ equals that of $(\Sigma)'$.*

Proof: By Theorems 3 and 4

$$l_{\{\Sigma\}}\Sigma = l_{(\{\Sigma\})'}(\Sigma)' = l_{\{\{\Sigma\}\}'}(\Sigma)' = l_{\{\Sigma\}'}(\Sigma)'.$$

5. A certain differential ideal

Let $\mathfrak{F}$ be any differential field of characteristic zero, and let y be an unknown. We work in the differential ring $\mathfrak{F}\{y\}$. The following theorem uses an idea due to Raudenbush.[15]

Theorem 5: *$[y]^2$ does not have a strong basis.*

Proof: Assume that $[y]^2$ has a strong basis. Then it has a strong basis consisting purely of forms $y_i y_j$. Let s be an integer so large that the set of forms

$$(7) \qquad y_i y_j, \qquad 0 \leqq i \leqq j \leqq s,$$

is a strong basis of $[y]^2$. Let p be the exponent with respect to $[y]^2$ of the differential ideal generated by the basis (7). Let α be a positive integer whose value (large) is to be given later.

[15] See paper cited under footnote 1.

Consider the forms

$$y_{i_1} y_{i_2} \cdots y_{i_{2p}}, \qquad i_1 + \cdots + i_{2p} = \alpha. \tag{8}$$

Each form (8) is in $[y]^{2p}$. Considering degree and weight, we see that each form (8) is a linear combination, with coefficients in $\mathfrak{F}$, of forms

$$(y_i y_j)_k y_{j_1} \cdots y_{j_{2p-2}},$$

$$0 \leqq i \leqq j \leqq s, \qquad i + j + k + j_1 + \cdots + j_{2p-2} = \alpha. \tag{9}$$

Hence the number of linearly independent forms (8) does not now exceed the number of forms (9). Since the totality of forms (8) are linearly independent, it follows that the number of distinct forms (8) is less than or equal to the number of forms (9).

Now, the number of expressions (8) is clearly at least

$$\frac{1}{(2p)!}\left(\frac{\alpha}{2p} + 1\right)^{2p-1}$$

if we suppose that α is divisible by $2p$. For, we may assign $i_1, \cdots, i_{2p-1}$ arbitrary values from 0 to $\alpha/2p$, and then i_{2p} is uniquely determined; moreover, the number of times a given form (8) can be produced in this way is at most $(2p)!$ On the other hand, the number of expressions (9) is surely not more than

$$\tfrac{1}{2}(s + 1)(s + 2)(\alpha + 1)^{2p-2},$$

for the number of forms in the basis (7) is $(s + 1)(s + 2)/2$. But if α is sufficiently large, then

$$\frac{1}{(2p)!}\left(\frac{\alpha}{2p} + 1\right)^{2p-1} > \tfrac{1}{2}(s + 1)(s + 2)(\alpha + 1)^{2p-2}.$$

Hence, the number of forms (8) exceeds the number of forms (9). This contradiction completes the proof.

COROLLARY 2: *Let $P \in \mathfrak{F}\{y\}$ be of order zero. Then the exponent of $[P]$ is ∞ or 1 according as P has or has not a multiple factor.*

PROOF: By Corollary 1 we may suppose that $\mathfrak{F}$ contains all the roots of $P = 0$. Let

$$P = b(y - \eta_1)^{k_1} \cdots (y - \eta_s)^{k_s}, \qquad \eta_i \neq \eta_j \text{ if } i \neq j,$$

be the factorization of P into linear factors. It is easy to see that

$$[P] = [(y - \eta_1)^{k_1}] \cdots [(y - \eta_s)^{k_s}] \tag{10}$$

is the product representation of $[P]$. If $k_i = 1$, then $[(y - \eta_i)^{k_i}] = [y - \eta_i]$ obviously has exponent unity. If $k_i > 1$, then $[(y - \eta_i)^{k_i}]$ has exponent ∞. For, otherwise $(y - \eta_i)^{k_i}$ would be a strong basis for $[y - \eta_i]^2$, a contradiction of Theorem 5. Since by Theorem 1 the exponent of $[P]$ equals the maximum of the exponents of the $[(y - \eta_i)^{k_i}]$, the proof is complete.

Part II. The Differential Ideal Generated by a Form in One Unknown of Order Unity

6. Preliminary remarks

Consider a form A in a single unknown y, of order unity. We seek the exponent of $[A]$. We shall work in a fixed differential field $\mathfrak{F}$ of characteristic zero, which contains the coefficients of A. By Corollary 1, we may (and do) assume, without further mention, that $\mathfrak{F}$ is extensive enough to contain certain solutions of A which we shall discuss.

The exponent of $[A]$ will be found to depend on the nature of the *singular solutions* of A, that is, the solutions of A which annul S, the separant of A.[16] A singular solution of A must annul the resultant with respect to y_1, the derivative of y, of A and S. This resultant is a nonzero form of order zero, or is 0, according as A has no multiple factor of order unity, or has such a factor, that is, according as A and S have not or have a common factor of order unity. In the former case, A has only a finite number of singular solutions, and these are easily found.

If $y = \eta$ is a singular solution of A, we consider A as a form in $z = y - \eta$, and write

$$A = \sum_{\nu=0}^{n} P_\nu z^{p_\nu} z_1^\nu, \tag{11}$$

where each P_ν either is of order zero and not divisible by z, or vanishes, and where the p_ν are definite non-negative integers.[17] We suppose that the degree of A in y_1 is n, so that $P_n \neq 0$.

Let m be the minimum of the total degrees of all terms effectively present in A considered as a polynomial in z and z_1. We shall call m the *multiplicity* of the singular solution $y = \eta$. It is obvious that if A has a multiple factor, any solution of that factor is a singular solution of multiplicity greater than unity.

7. Singular solutions of multiplicity exceeding unity

We shall prove that *if A has a singular solution of multiplicity exceeding unity, then the exponent of $[A]$ is ∞.*

(a) Let $y = \eta$ be a singular solution of A of multiplicity at least 2, and suppose that $y = \eta$ is an essential manifold of A. Let $z = y - \eta$. The product representation of $[A]$ must have a factor whose manifold is $y = \eta$, that is, a factor Σ_1 such that $\{\Sigma_1\} = \{z\}$. Now, it is known that

$$\Sigma_1 = [A, z^q], \tag{12}$$

[16] The definition here of *singular solution* is broader than that given by Ritt in his book, *Differential equations from the algebraic standpoint*, and used in the Appendix of the present paper, for A is not assumed here to be algebraically irreducible.

[17] If $P_\nu = 0$, we let p_ν be any fixed non-negative integer.

where q is any sufficiently large integer.[18] Because $y = \eta$ is of multiplicity at least 2, it follows that $A \epsilon [z^2, z_1^2]$. Hence, by (12),

$$\Sigma_1 \subseteq [z^2, z_1^2].$$

Thus, the exponent of Σ_1 is not less than that of $[z^2, z_1^2]$, which, by Theorem 5, is ∞. Therefore, using Theorem 1, we see that the exponent of $[A]$ is ∞.

(b) Suppose that $y = \eta$ is a singular solution of A of multiplicity greater than unity, which is not an essential manifold of A. Let $y - \eta = z$. We shall show first that $\{A\}$ contains a form which, when considered as a form in z, has at least one term of the first degree.

Let $C_1, \cdots, C_q$ be the distinct irreducible factors of A of order unity, and let $F = C_1 \cdots C_q$. If T denotes the separant of F, the derivative of F is a form $Tz_2 + U$, where U is a form of order not exceeding unity. Hence

$$Tz_2 \equiv -U \qquad [F].$$

It follows, by successive differentiations and eliminations, that

$$T^{a_i} z_i \equiv V_i \qquad [F], \qquad\qquad i = 2, \cdots, n + 2,$$

where the a_i are appropriate integers, and the V_i are forms whose order is not greater than unity. Letting a be the maximum of the a_i, we have

$$T^a z_i \equiv W_i \qquad [F], \qquad\qquad i = 2, \cdots, n + 2,$$

where $W_i = T^{a-a_i} V_i$.

Let J be the initial of F. Then J is of order zero. Clearly, there exists a non-negative integer b such that

$$J^b W_i \equiv X_i \qquad (F), \qquad\qquad i = 2, \cdots, n + 2,$$

where each X_i is of order not exceeding unity, and is of degree less than n in z_1.[19] It follows that

$$J^b T^a z_i \equiv X_i \qquad [F], \qquad\qquad i = 2, \cdots, n + 2.$$

The $n + 1$ forms X_i are linear combinations, with coefficients which are polynomials in z, of the n quantities $1, z_1, \cdots, z_1^{n-1}$. Hence, some linear combination of the X_i, with coefficients which are forms of order zero, not all 0, must vanish:

$$\sum_{i=2}^{n+2} K_i X_i = 0.$$

We assume, as we may, that the K_i are not all divisible by z. Referring to the last congruence, we now see that

[18] Ritt and Kolchin, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 895–898.

[19] The degree of F in z_1 is at most n.

$$J^bT^aM \equiv 0 \qquad [F],$$

where $M = \sum_{i=2}^{n+2} K_i z_i$.

Since not every K_i is divisible by z, M has at least one term of degree unity. Consider the form $y - \zeta$, where ζ is either a common solution of TJ and F, or a root of a factor of A of order zero, but is distinct from η. Let N be the product of all such forms $y - \zeta$, and let $Z = MN$.

Z, like M, has at least one term of degree unity (when considered as a form in z). Z vanishes for all solutions of F with the possible exception of $y = \eta$, and for all the solutions of A which are not solutions of F, again with the possible exception of $y = \eta$. That is, Z vanishes for all solutions of A, save possibly $y = \eta$. Since $y = \eta$ is not an essential manifold, Z must also vanish for $y = \eta$. By the analog to the Hilbert-Netto theorem,[20] then, $Z \epsilon \{A\}$. Z is the form whose existence we were to prove.

Thus, $\{A\}$ contains a form $z_j + H$, where H, considered as a form in z, has each of its terms either of order less than j or of degree greater than unity.

Assume, now, that the exponent of $[A]$ is finite, say q. Then[21]

$$(z_{j+i_1} + H_{i_1}) \cdots (z_{j+i_q} + H_{i_q}) \equiv 0 \qquad [A]$$

for all non-negative integers $i_1, \cdots, i_q$. Since $y = \eta$ is of multiplicity at least 2, as a singular solution of A, it follows that $A \epsilon [z^2, z_1^2]$. Hence

$$(z_{j+i_1} + H_{i_1}) \cdots (z_{j+i_q} + H_{i_q}) \equiv 0_i \qquad [z^2, z_1^2].$$

If we compare terms here of degree q and weight $qj + i_1 + \cdots + i_q$, we obtain

$$z_{j+i_1} \cdots z_{j+i_q} \equiv 0 \qquad [z^2, z_1^2], \qquad\qquad i_1, \cdots, i_q = 0, 1, 2, \cdots.$$

This is easily seen to imply, however, that $[z]^2$ has a strong basis. This contradiction of Theorem 5 completes the proof.

EXAMPLE 1. Let $A = \sum_{\nu=0}^{n} Q_\nu y_1^\nu$ $(n \geqq 1)$, where the Q_ν are forms of order zero with *constant* coefficients, and where $Q_n \neq 0$.

If A has no multiple factors of order unity, the singular solutions must be constants, for they must satisfy an algebraic equation with constant coefficients;[22] since $S = \sum_{\nu=1}^{n} \nu Q_\nu y_1^{\nu-1}$, it is evident that the singular solutions of A are precisely the common zeros of Q_0 and Q_1. A singular solution is of multiplicity exceeding unity if it is a multiple zero of Q_0. Thus, *if A has a multiple factor, or if Q_0 has a multiple factor which is also a factor of Q_1, then the exponent of $[A]$ is ∞.*

8. Singular solutions of multiplicity unity; the product representation

Throughout §§8–10 we suppose that A has singular solutions, and that they are all of multiplicity unity. In the present section we shall show that all the

[20] See the paper cited under footnote 2.

[21] H_{i_k} is the i_k-th derivative of H.

[22] The equation obtained by eliminating z_1 from $A = 0$, $S = 0$.

singular solutions are essential manifolds, and shall find the product representation of $[A]$. With an additional assumption, we shall determine the exponent of $[A]$ (§10).

Since A has no singular solution of multiplicity exceeding unity, A has no multiple factors. Hence A has only a finite number of singular solutions (§6).

Let $y = \eta$ be a singular solution of A, and write $z = y - \eta$. Since S must vanish for $y = \eta$, we see from (11) that either $P_1 = 0$ or $p_1 > 0$. This means, since $y = \eta$ is of multiplicity unity, that $P_0 \neq 0$ and $p_0 = 1$. (11) thus becomes

$$A = P_0 z + \sum_{\nu=1}^{n} P_\nu z^{p_\nu} z_1^\nu. \tag{13}$$

It now follows from a theorem of Ritt that $y = \eta$ is an essential manifold of A.[23]

We now turn to the product representation of $[A]$.

Let $y = \eta_i$, $i = 1, \cdots, s$, be the singular solutions of A; we denote $y - \eta_i$ by u_i. It is evident that we may write

$$[A] = \Sigma_0 \Sigma_1 \cdots \Sigma_s, \tag{14}$$

where the manifold of Σ_0 consists of the non-singular solutions of A,[24] and where the manifold of Σ_i, $i > 0$, is $y = \eta_i$, that is, $\{\Sigma_i\} = \{u_i\}$, $i = 1, \cdots, s$.

Referring to (13) and making use of a known result,[25] we see that

$$\Sigma_i = [u_i], \qquad i = 1, \cdots, s. \tag{15}$$

The determination of Σ_0 is more difficult. The first step will be to find $\{\Sigma_0\}$.

If z is one of the u_i, then (13) holds. Differentiating (13) we obtain

$$A_1 = P_{01} z + \sum_{\nu=1}^{n} P_{\nu 1} z^{p_\nu} z_1^\nu + z_1 \left(P_0 + \sum_{\nu=1}^{n} p_\nu P_\nu z^{p_\nu - 1} z_1^\nu\right) + \sum_{\nu=1}^{n} \nu P_\nu z^{p_\nu} z_1^{\nu-1} \cdot z_2,$$

where $P_{\nu 1}$ is the derivative of P_ν. But by (13),

$$P_0 z \equiv -z_1 \sum_{\nu=1}^{n} P_\nu z^{p_\nu} z_1^{\nu-1} \qquad (A). \tag{16}$$

Hence

$$P_0 A_1 = P_{01} P_0 z + P_0 z_1 \sum_{\nu=1}^{n} P_{\nu 1} z^{p_\nu} z_1^{\nu-1} + P_0 z_1 \left(P_0 + \sum_{\nu=1}^{n} p_\nu P_\nu z^{p_\nu-1} z_1^\nu\right)$$
$$+ \left(P_1 z^{p_1-1} \cdot P_0 z + P_0 z_1 \sum_{\nu=2}^{n} \nu P_\nu z^{p_\nu} z_1^{\nu-2}\right) z_2$$

---

[23] Ritt, Annals of Mathematics, vol. 37 (1936), pp. 552–617. See especially §5. Professor Ritt's proof is not applicable to abstract differential fields, as it involves function-theoretic concepts. But an abstract proof of the *sufficiency* part of the theorem (we need only the sufficiency) has recently been achieved by H. Levi, and may be expected to appear in the literature shortly.

[24] If A is algebraically irreducible, the manifold of Σ_0 is the general solution of A.

[25] See §2 of the paper cited under footnote 18.

$$\equiv z_1\Big(-P_{01}\sum_{\nu=1}^{n} P_\nu z^{p_\nu} z_1^{\nu-1} + P_0 \sum_{\nu=1}^{n} P_{\nu 1} z^{p_\nu} z_1^{\nu-1} + P_0\Big(P_0 + \sum_{\nu=1}^{n} p_\nu P_\nu z^{p_\nu -1} z_1^{\nu}\Big)$$

$$+\Big(-P_1 z^{p_1-1}\sum_{\nu=1}^{n} P_\nu z^{p_\nu} z_1^{\nu-1} + P_0 \sum_{\nu=2}^{n} \nu P_\nu z^{p_\nu} z_1^{\nu-2}\Big) z_2\Big), \qquad (A),$$

so that

$$P_0 A_1 \equiv z_1(U z_2 + V) \qquad (A), \tag{17}$$

where

$$\begin{aligned} U &= -P_1 z^{p_1-1}\sum_{\nu=1}^{n} P_\nu z^{p_\nu} z_1^{\nu-1} + P_0 \sum_{\nu=2}^{n} \nu P_\nu z^{p_\nu} z_1^{\nu-2}, \\ V &= -P_{01}\sum_{\nu=1}^{n} P_\nu z^{p_\nu} z_1^{\nu-1} + P_0 \sum_{\nu=1}^{n} P_{\nu 1} z^{p_\nu} z_1^{\nu-1} + P_0\Big(P_0 + \sum_{\nu=1}^{n} p_\nu P_\nu z^{p_\nu-1} z_1^{\nu}\Big). \end{aligned} \tag{18}$$

From (17) and (16) we find that

$$z(P_0 U z_2 + P_0 V) \equiv 0 \qquad (A, A_1).$$

This implies that

$$\{A\} = \{z\} \cap \{A, P_0 U z_2 + P_0 V\}.$$

The two ideals in the second member here are separated, for by (18)

$$P_0 V \equiv P_0^3 \not\equiv 0 \qquad [z], \tag{19}$$

so that $P_0 U z_2 + P_0 V \not\equiv 0 \;\; [z]$.

Now, z represents any one of the u_i. Let

$$I_i = P_0 U, \qquad J_i = P_0 V \quad \text{when} \quad z = u_i, \qquad i = 1, \cdots, s. \tag{20}$$

Then by the above, for $i = 1, \cdots, s$,

$$\{A\} = \{u_i\} \cap \{A, I_i u_{i2} + J_i\},$$

$$u_i(I_i u_{i2} + J_i) \equiv 0, \qquad u_{i1}(I_i u_{i2} + J_i) \equiv 0 \qquad (A, A_1), \tag{21}$$

and $y = \eta_i$ is not a solution of $I_i u_{i2} + J_i$. It readily follows that

$$\{A\} = \{A, I_1 u_{12} + J_1, \cdots, I_s u_{s2} + J_s\} \cap \{u_1\} \cap \cdots \cap \{u_s\}. \tag{22}$$

The $s + 1$ differential ideals, whose intersection appears in the second member of (23), are separated, so that we may replace intersection by product.

Comparing (22) with

$$\{A\} = \{\Sigma_0\} \cap \{\Sigma_1\} \cap \cdots \cap \{\Sigma_s\}$$

(which follows from (14)) and with (15), we see that

$$\{\Sigma_0\} = \{A, I_1 u_{12} + J_1, \cdots, I_s u_{s2} + J_s\}. \tag{23}$$

We shall now obtain two forms M and N such that

$$(24) \qquad M \epsilon \{u_1\} \cdots \{u_s\}, \qquad N \epsilon \{\Sigma_0\}, \qquad MN \epsilon [A], \qquad M + N = 1.$$

This will enable us to determine Σ_0.

Let

$$B_i = I_i u_{i2} + J_i, \qquad i = 1, \cdots, s,$$

and let

$$Q_i = P_0 \quad \text{when} \quad z = u_i, \qquad i = 1, \cdots, s.$$

Then Q_i is not divisible by u_i, and, by (19) and (20),

$$(25) \qquad B_i - Q_i^3 \equiv I_i u_{i2} \qquad (u_i, u_{i1}).$$

Since Q_i is not divisible by u_i, there is a form C_i of order zero, and an element $a_i \epsilon \mathfrak{F}$, such that

$$1 = a_i Q_i^3 + C_i u_i.^{26}$$

This may be written

$$1 = [C_i u_i - a_i(B_i - Q_i^3)] + a_i B_i.$$

By (21), (25) and the definition of B_i, moreover,

$$[C_i u_i - a_i(B_i - Q_i^3)] \cdot a_i B_i \equiv 0 \qquad [A],$$

provided $I_i \equiv 0 \ (u_i, u_{i1})$, that is, provided $P_2 = 0$ or $p_2 > 0$ when $z = u_i$ (we refer to (20) and (18), and the fact that $p_1 > 0$). If this is not the case, then

$$[C_i u_i - a_i(B_i - Q_i^3)] \cdot (a_i B_i)^2 \equiv 0 \qquad [A],$$

because $u_{i2} B_i^2 \equiv 0$ is a consequence of $u_{i1} B_i \equiv 0$, and the latter congruence holds by (21).

Accordingly, we separate the singular solutions of A into two classes. A singular solution $y = \eta$ of A will be said to be of the *first class* if in A, considered as a form in $z = y - \eta$, the coefficient of z_1^2 is divisible by z; a singular solution which is not of the first class will be said to be of the *second class.*

We suppose the singular solutions η_i ordered so that $\eta_1, \cdots, \eta_r$ are of the first class, and $\eta_{r+1}, \cdots, \eta_s$ are of the second class. This is the same as supposing that

$$(26) \qquad \begin{aligned} &P_2 = 0 \quad \text{or} \quad p_2 > 0 \quad \text{when} \quad z = u_i, && i = 1, \cdots, r, \\ &P_2 \neq 0 \quad \text{and} \quad p_2 = 0 \quad \text{when} \quad z = u_i, && i = r + 1, \cdots, s. \end{aligned}$$

We let

$$(27) \qquad \begin{aligned} M_i &= C_i u_i - a_i(B_i - Q_i^3), && i = 1, \cdots, r, \\ M_i &= [C_i u_i - a_i(B_i - Q_i^3)]^2 \\ &\qquad + 2[C_i u_i - a_i(B_i - Q_i^3)] \cdot a_i B_i, && i = r + 1, \cdots, s, \\ N_i &= a_i B_i, && i = 1, \cdots, r, \\ N_i &= (a_i B_i)^2, && i = r + 1, \cdots, s. \end{aligned}$$

[26] It will be observed that $a_i \neq 0$.

Then $M_i \epsilon \{u_i\}$, $N_i \epsilon \{\Sigma_0\}$, $M_i + N_i = 1$, $M_iN_i \epsilon [A]$, $i = 1, \cdots, s$. Letting, therefore,

$$(28)\qquad \begin{aligned} M &= M_1 \cdots M_s \\ N &= (M_1 + N_1) \cdots (M_s + N_s) - M_1 \cdots M_s = 1 - M, \end{aligned}$$

we see that (24) holds.

We refer now to the paper cited under footnote 5. The proof in that paper shows that $\Sigma_0 = [A, N]$. But, by (28), $N \equiv 0\ (N_i, \cdots, N_s)$, and $N_i = MN_i + NN_i \equiv 0\ [A, N]$. Hence $\Sigma_0 = [A, N_1, \cdots, N_s]$. Referring to (27), we see, finally, that

$$(29)\qquad \Sigma_0 = [A, B_1, \cdots, B_r, B_{r+1}^2, \cdots, B_s^2].$$

The product representation of A is given by (14), (15) *and* (29).[27]

9. Continuation

The purpose of this section is to prove that

$$(30)\qquad [A, B_1, \cdots, B_s]^2 \subseteq \Sigma_0 .$$

By (29) it suffices to show that

$$(31)\qquad B_{\sigma i}B_{\tau j} \equiv 0\ \ (\Sigma_0), \qquad \sigma, \tau = 1, \cdots, s, \qquad i, j = 0, 1, 2, \cdots .$$

Let $Iz_2 + J$ represent any $B_\sigma = I_\sigma u_{\sigma 2} + J_\sigma$. We shall prove that

$$(32)\qquad (Iz_2 + J)_i(Iz_2 + J)_j \equiv 0 \qquad (\Sigma_0)$$

for all i, j.

From (21) we see that

$$(33)\qquad z(Iz_2 + J) \equiv 0, \qquad z_1(Iz_2 + J) \equiv 0 \qquad [A].$$

Differentiating the second congruence here $l + 1$ times, we find that

$$(34)\qquad \sum_{\lambda=0}^{l+1} \binom{l+1}{\lambda} z_{\lambda+1}(Iz_2 + J)_{l+1-\lambda} \equiv 0 \qquad [A].$$

By (29), (32) holds for $i = 0, j = 0$. We assume that (32) holds for $i = 0, 1, \cdots, l - 1$ and $j = 0$, and prove it for $i = l, j = 0$.

Multiplying (34) by $Iz_2 + J$, we obtain, using (33) and the induction assumption,

$$(l + 1)z_2(Iz_2 + J)_l(Iz_2 + J) \equiv 0 \qquad (\Sigma_0).$$

But $Iz_2(Iz_2 + J) \equiv - J(Iz_2 + J)$, (Σ_0). Hence

$$J(Iz_2 + J)_2(Iz_2 + J) \equiv 0 \qquad (\Sigma_0).$$

[27] Actually, Σ_0 may not be connected. This will occur if A has more than one factor of order unity.

However, by (20) and (19), J is congruent, modulo (z, z_1), to a nonzero element of $\mathfrak{F}$. Hence, by (33), $(Iz_2 + J)_i(Iz_2 + J) \equiv 0 \ (\Sigma_0)$.

Thus (32) holds for all i when $j = 0$.

To prove (32) for all i and all j we again use induction. Assume (32) to be valid for all i and for $j = l - 1$. Differentiating (32) for $j = l - 1$ we see that

$$(Iz_2 + J)_{i+1}(Iz_2 + J)_{l-1} + (Iz_2 + J)_i(Iz_2 + J)_l \equiv 0 \qquad (\Sigma_0),$$

so that, by the induction assumption, $(Iz_2 + J)_i(Iz_2 + J)_l \equiv 0 \ (\Sigma_0)$. This proves (32) for all i and j.

Thus (31) holds whenever $\sigma = \tau$. To prove (31) in the general case, consider $B_{\sigma i}B_{\tau j}$. It is sufficient, of course, to suppose that $\sigma > \tau$, for $B_\sigma \in \Sigma_0$ if $\sigma \leqq \tau$. Now, the remainder of $B_{\tau j}$ with respect to B_σ is a form in $\{\Sigma_0\}$ of order not exceeding unity. Such a form must be divisible by A. Hence, for an appropriate integer m, $I_\sigma^m B_{\tau j} \equiv 0 \ [A, B_\sigma]$. Hence, since (31) holds when $\tau = \sigma$, $I_\sigma^m B_{\sigma i}B_{\tau j} \equiv 0 \ (\Sigma_0)$. But $\sigma > \tau$, so that (see (26), (20) and (18)) I_σ is congruent modulo $(u_\sigma, u_{\sigma 1})$, to a nonzero element of $\mathfrak{F}$. Hence, by (21), $B_{\sigma i}B_{\tau j} \equiv 0 \ (\Sigma_0)$.

10. The regular type

We shall say that A is of the *regular type* if

$$(35) \qquad 1 \equiv 0 \ (\Sigma_0, S, I_1, \cdots, I_s).$$

In the present section *we assume that A is of the regular type.*

We shall prove that *the exponent of $[A]$ is 1 or 2 according as all of the singular solutions of A are of the first class or at least one singular solution of A is of the second class.*

(a) Let all the singular solutions of A be of the first class, and let $G \in \{\Sigma_0\}$. We shall show that $G \in \Sigma_0$.

For each i we have, for an appropriate integer b_i, $I_i^{b_i}G \equiv G_i'[B_i]$, where G_i' is of order unity, at most. Since $G_i' \in \{\Sigma_0\}$, G_i' must be divisible by A. Hence, for $i = 1, \cdots, s$, $I^{b_i}G \equiv 0 \ [A, B_i]$. Similarly, for a suitable c, $S^cG \equiv 0 \ [A]$. Now, by (35), there are forms X_i, and a form $T \in \Sigma_0$, such that

$$1 = T + X_0S^c + X_1I_1^{b_1} + \cdots + X_sI_s^{b_s}.$$

Therefore,

$$G = (T + X_0S^c + X_1I_1^{b_1} + \cdots + X_sI_s^{b_s})G \equiv 0 \qquad [A, B_1, \cdots, B_s].$$

But all the singular solutions of A are of the first class. Therefore, by (29), $\Sigma_0 = [A, B_1, \cdots, B_s]$. Hence, $G \in \Sigma_0$.

Thus, the exponent of Σ_0 is unity. But the exponent of $\Sigma_i = [u_i]$, $i = 1, \cdots, s$, is obviously unity. Hence, by Theorem 1, unity is also the exponent of $[A]$.

(b) Let A have at least one singular solution of the second class, and let $G \in \{\Sigma_0\}$. As in (a) we see that $G \in [A, B_1, \cdots, B_s]$. This, together with (31), proves that the exponent of Σ_0 does not exceed 2. Hence the exponent of $[A]$ is no greater than 2.

To prove that the exponent of $[A]$ is exactly 2, it suffices to exhibit a form in $\{A\}$ which is not in $[A]$. To this end, we let z be any u_i with $i > r$, and show that $\{\Sigma_0\}$ contains a form $z_h + H$, where h is a sufficiently large positive integer, and H is a form, of order less than h, contained in $[z]$.

By (35) there is a relation

$$1 = T + X_0S + X_1I_1 + \cdots + X_sI_s, \qquad T \in \Sigma_0. \tag{36}$$

Let $h(\geqq 3)$ be an integer such that $h - 1$ exceeds the orders of $T, X_0, \cdots, X_s$. If we differentiate each B_i, written as a form in z, $h - 3$ times, we obtain relations $I_iz_{h-1} + K_i \equiv 0\ \{\Sigma_0\}$. Similarly, differentiating A $h - 2$ times we obtain $Sz_{h-1} + K_0 \equiv 0\ \{\Sigma_0\}$. All K_i are forms of order less than $h - 1$. These relations, together with $Tz_{h-1} \equiv 0\{\Sigma_0\}$, yield, because of (36), a relation $z_{h-1} + K \equiv 0\ \{\Sigma_0\}$, where K is a form of order less than $h - 1$.

Let α be the term of K free of z and the derivatives of z. If $\alpha = 0$, then $(z_{h-1} + K)_1$ is a form $z_h + H$ as described above; if $\alpha \neq 0$, then $\alpha(\alpha^{-1}(z_{h-1} + K))_1$ is such a form.

Now let E be the product of all the u_i other than z, and consider $E(z_h + H)$, where $z_h + H$ is the form whose existence we have just proved. We see that $\{A\}$ contains a form $Ez_h + F$, where E is a form of order zero, not divisible by z, and F is a form of order less than h.

We shall now complete the proof by showing that $Ez_h + F \not\equiv 0\ [A]$. The method will be to assume that $Ez_h + F \equiv 0\ [A]$, and to force a contradiction.

Let then

$$Ez_h + F = C_0A + C_1A_1 + \cdots + C_tA_t, \tag{37}$$

where A_j represents the j-th derivative of A. By (17), for all $i > 0$,

$$P_0A_i \equiv (z_1(Uz_2 + V))_{i-1} \qquad (A, A_1, \cdots, A_{i-1}).$$

It follows from (37), therefore, that

$$\begin{aligned} P_0^t(Ez_h + F) = D_0A &+ D_1z_1(Uz_2 + V) \\ &+ D_2(z_1(Uz_2 + V)_1 + z_2(Uz_2 + V)) \\ &+ \cdots + D_t \sum_{i=0}^{t-1} \binom{t-1}{i} z_{i+1}(Uz_2 + V)_{t-1-i}. \end{aligned} \tag{38}$$

We now consider the various forms in (38) as polynomials in the z_i. We shall modify (38) in such a way as to make evident a contradiction

Consider a power series[28] $\sum_{i=1}^{\infty} \beta_i z_1^i$ in z_1 such that A vanishes when $\Sigma\beta_i z_1^i$ is substituted for z. Such a power series exists, as by (13) $z_1 = 0, z = 0$ renders A zero and leaves $\dfrac{\partial A}{\partial z}$ nonzero.[29]

[28] All power series which we mention are formal.

[29] The "implicit function theorem" for abstract fields of characteristic zero may be proved by the method of undetermined coefficients. The proof is simpler than the one in analysis, because the series is formal and no questions of convergence arise.

Substituting $\Sigma\beta_i z_1^i$ for z in (38), we find a relation

$$
(39)\qquad \begin{aligned} E'z_h + F' = D_1' z_1(U'z_2 + W_0) + D_2'(z_1(U'z_3 + W_1) + z_2(U'z_2 + W_0)) \\ + \cdots + D_t' \sum_{i=0}^{t-1} \binom{t-1}{i} z_{i+1}(U'z_{t+1-i} + W_{t-1-i}). \end{aligned}
$$

Here E', F', U', the D_i' and the W_i are power series in z_1 with coefficients which are polynomials in the z_i with $i > 1$. Each W_i is free of z_j with $j > i + 1$. Moreover, E' and U' involve only z_1, and have nonzero terms of degree zero. For E' this is true because $P_0^l E$ is not divisible by z; for U' because $P_2 \neq 0$ and $p_2 = 0$ (z is a u_i with $i > r$; see (26) and (18).)

Let

$$
(40)\qquad w_1 = z_1, \quad w_i = U'z_i + W_{i-2}, \qquad i = 2, 3, \cdots.
$$

Since U' has a nonzero term of degree zero, the equations (40) may be inverted to obtain

$$
(41)\qquad z_1 = w_1, \quad z_i = Yw_i + Z_i, \qquad i = 2, 3, \cdots,
$$

where Y is a series in w_1 with a nonzero term of degree zero, and each Z_i is a series in w_1 with coefficients which are polynomials in $w_2, \cdots, w_{i-1}$. The equations (41) define a substitution, with inverse substitution (40). Applying substitution (41) to (39), we obtain a relation

$$
(42)\qquad \begin{aligned} Qw_h + R = L_1 w_1 w_2 + L_2(w_1 w_3 + (Yw_2 + Z_2)w_2) \\ + \cdots + L_t(w_1 w_{t+1} + (t-1)(Yw_2 + Z_2)w_t + \cdots + (Yw_t + Z_t)w_2), \end{aligned}
$$

where Q is a series in w_1 with a nonzero term a of degree zero, and R is free of the w_i with $i \geqq h$.

For the second member of (42) to produce the term aw_h which appears in the first member, some L_i with $i \geqq h$ must have a nonzero term of degree zero, because each term in the coefficient of L_i is of at least the first degree in $w_1, w_2, \cdots, w_i$. Thus, some of the L_i have a term which is simply an element of $\mathfrak{F}$, and hence certainly terms bw_1^j ($b \in \mathfrak{F}$, $b \neq 0$, $j \geqq 0$). Let L_p be the L_i of highest subscript, with this property. Then $p \geqq h$. Of all the terms of L_p of the above mentioned type, let cw_1^k be the one of least degree. Then

$$
L_p(w_1 w_{p+1} + \cdots + (Yw_p + Z_p)w_2)
$$

contains the term $cw_1^{k+1}w_{p+1}$. This term does not appear in the first member of (42), and must therefore be cancelled by terms in some expression

$$
(43)\qquad L_i(w_1 w_{i+1} + \cdots + (Yw_i + Z_i)w_2)
$$

with $i \neq p$. But if $i < p$, then each term of (43) is divisible by some w_j with $2 \leqq j \leqq p$. Hence $cw_1^{k+1}w_{p+1}$ must be cancelled by terms from expressions (43) with $i > p$. This implies that some L_i with $i > p$ has a term of the type bw_1^{k+1} ($b \in \mathfrak{F}$, $b \neq 0$). This contradicts the definition of L_p, and completes the proof.

The following example shows that of the forms A with *constant coefficients*, which have singular solutions, all of multiplicity unity, those which are not of the regular type are exceptional. It also shows that, for any preassigned pair of integers r, s $(0 \leqq r \leqq s,\ s > 0)$, there exist forms of the regular type having r singular solutions of the first class and $s - r$ singular solutions of the second class.

EXAMPLE 2. Let r, s, m, n be integers with $0 \leqq r \leqq s \leqq m$, $s > 0$, $n \geqq 3$, let $u_i = y - \eta_i$, $i = 1, \cdots, s$, where $\eta_1, \cdots, \eta_s$ are distinct constant elements of a differential field $\mathfrak{E}$ of characteristic zero, and let

$$Q_0 = \sum_{\mu=0}^{m-s} a_{\mu 0} y^\mu, \qquad Q_1 = \sum_{\mu=0}^{m-s} a_{\mu 1} y^\mu, \qquad Q_2 = \sum_{\mu=0}^{m-r} a_{\mu 2} y^\mu,$$

$$Q_\nu = \sum_{\mu=0}^{m} a_{\mu\nu} y^\mu, \qquad \nu = 3, \cdots, n,$$

where $a_{00}, a_{01}, \cdots, a_{mn}$ is a set of constants, algebraically independent with respect to $\mathfrak{E}$. Consider the form

$$A = Q_0 u_1 \cdots u_s + Q_1 u_1 \cdots u_s y_1 + Q_2 u_1 \cdots u_r y_1^2 + \sum_{\nu=3}^{n} Q_\nu y_1^\nu$$

with coefficients in $\mathfrak{F} = \mathfrak{E}\langle a_{00}, a_{01}, \cdots, a_{mn}\rangle$.

We shall show that the singular solutions of A are $\eta_1, \cdots, \eta_s$, where $\eta_1, \cdots, \eta_r$ are of the first class, and $\eta_{r+1}, \cdots, \eta_s$ are of the second. We shall prove, moreover, that A is of the regular type. In fact, we shall prove the stronger result that

$$(44) \qquad 1 \equiv 0 \quad (A, S, I_1, \cdots, I_s, J_1, \cdots, J_s).$$

Let R be the resultant with respect to y_1 of A and S. R is a polynomial in y and the $a_{\mu\nu}$. If we let the $a_{\mu\nu}$ take on special values, such that A becomes $u_1 \cdots u_s + y_1^n$, then S becomes ny_1^{n-1}, so that R becomes $n^n(u_1 \cdots u_s)^{n-1} \neq 0$. Since R is not 0 for special values of the $a_{\mu\nu}$, R does not vanish for unknown $a_{\mu\nu}$.

Now, each singular solution is a solution of R, which is of order zero and has constant coefficients. Since $R \neq 0$, this implies that all the singular solutions of A are constants. Hence, from the expression for A and the expression for S derived therefrom, we see that a singular solution of A must annul $Q_0 u_1 \cdots u_s$ and $Q_1 u_1 \cdots u_s$, that is, must be one of the η_i. That each η_i is a singular solution is obvious. Thus, the singular solutions of A are $\eta_1, \cdots, \eta_s$. It is apparent that these are of multiplicity unity, and that η_i is of the first or second class according as $i \leqq r$ or $i > r$.

We find, using (20) and (18) that

$$I_i \equiv Q_0(u_1 \cdots u_{i-1}u_{i+1} \cdots u_s)^2 I, \qquad J_i \equiv (Q_0 u_1 \cdots u_{i-1}u_{i+1} \cdots u_s)^2 J, \quad (A),$$

where

$$I = -Q_1^2 u_1 \cdots u_s - Q_1 Q_2 u_1 \cdots u_r y_1 - Q_1 \sum_{\nu=3}^{n} Q_\nu y_1^{\nu-1} + 2Q_0 Q_2 u_1 \cdots u_r + Q_0 \sum_{\nu=3}^{n} \nu Q_\nu y_1^{\nu-2},$$

$$J = (Q_0 u_1 \cdots u_s)' + (Q_1 u_1 \cdots u_s)' y_1 + (Q_2 u_1 \cdots u_r)' y_1^2 + \sum_{\nu=3}^{n} Q_\nu' y_1^\nu,$$

accents indicating differentiation with respect to y. Hence,

$$(A, S, I_1, \cdots, I_s, J_1, \cdots, J_s) = (A, S, Q_0 I, Q_0^2 J).$$

Suppose, now, that (44) is false. Then there is a pair of values $y = \bar{y}$, $y_1 = \bar{y}_1$ which annul the polynomials A, S, $Q_0 I$, $Q_0^2 J$. When $y_1 = 0$, A, S and $Q_0^2 J$ become, respectively, $Q_0 u_1 \cdots u_s$, $Q_1 u_1 \cdots u_s$ and $Q_0^2 (Q_0 u_1 \cdots u_s)'$, and these polynomials have no solution in common. Hence $\bar{y}_1 \neq 0$. Also, the resultant with respect to y_1 of S and

$$Q_1 u_1 \cdots u_s + Q_2 u_1 \cdots u_r y_1 + \sum_{\nu=3}^{n} Q_\nu y_1^{\nu-1} = \frac{A - Q_0 u_1 \cdots u_s}{y_1}$$

is a nonzero polynomial in y, relatively prime to Q_0.[30] Hence $\bar{y}$ is not a root of $Q_0 = 0$, so that $\bar{y}$, $\bar{y}_1$ annul I and J.

Let M be the resultant with respect to y_1, N the resultant with respect to y, of J and

$$nA - y_1 S = nQ_0 u_1 \cdots u_s + (n-1) Q_1 u_1 \cdots u_s y_1 + (n-2) Q_2 u_1 \cdots u_r y_1^2 + \sum_{\nu=3}^{n-1} (n-\nu) Q_\nu y_1^\nu.$$

M is a polynomial in y, N is a polynomial in y_1, each with coefficients which are polynomials in the $a_{\mu\nu}$ *not including* a_{0n}. Moreover, M and N are not 0.[31] Now, $\bar{y}$, $\bar{y}_1$ annul M and N. Thus, $\bar{y}$ and $\bar{y}_1$ are algebraic functions of the $a_{\mu\nu}$ other than a_{0n}. Since A vanishes for $y = \bar{y}$, $y_1 = \bar{y}_1$, and since $\bar{y}_1 \neq 0$, it follows that a_{0n} is an algebraic function of the other $a_{\mu\nu}$. This contradiction completes the proof.

---

[30] This resultant is nonzero because it is nonzero when $A = y_1^n + u_1 \cdots u_s y_1$; it is prime to Q_0 because its coefficients are independent of $a_{00}, \cdots, a_{m-s,0}$.

[31] If we specialize some of the $a_{\mu\nu}$ so that $A = y_1^n + y_1^{n-1} + Q_0 u_1 \cdots u_s$, then $J = (Q_0 u_1 \cdots u_s)'$, $nA - y_1 S = nQ_0 u_1 \cdots u_s + y_1^n$, and $M = (Q_0 u_1 \cdots u_s)'^{n-1} \neq 0$, so that M does not vanish before we specialize. If we specialize A further by setting $y_1 = 0$, then $J = (Q_0 u_1 \cdots u_s)'$, $nA - y_1 S = nQ_0 u_1 \cdots u_s$, and N is the resultant of $nQ_0 u_1 \cdots u_s$ and $(Q_0 u_1 \cdots u_s)^1$, which is not 0. Thus, N is not 0 before the specialization.

EXAMPLE 3. Let $n \geqq 2$, and let $A = Qy_1^n + P$, where P and Q are relatively prime nonzero forms of order zero with constant coefficients, P being of positive degree and without multiple factors. The singular solutions of A are the roots of $P = 0$, and are all of multiplicity unity.

If $z = y - \eta$, where η is a root of $P = 0$, then $\eta_1 = 0$, and we may write $A = (P/z)z + Qz_1^n$. Thus, referring to (18), we see that

$$U = n(P/z)Qz_1^{n-2},$$
$$V = -(P/z)'Qz_1^n + (P/z)Q'z_1^n + (P/z)^2.$$

Hence, if $\eta_1, \cdots, \eta_s$ are the roots of $P = 0$, then (see (20))

$$I_i = n\left(\frac{P}{y-\eta_i}\right)^2 Qy_1^{n-2},$$
$$J_i = \frac{P}{y-\eta_i}\left(-\left(\frac{P}{y-\eta_i}\right)' Qy_1^n + \frac{P}{y-\eta_i}Q'y_1^n\right) + \left(\frac{P}{y-\eta_i}\right)^3.$$

Since the roots of $P = 0$ are distinct, it follows that $Qy_1^{n-2} \equiv 0(I_1, \cdots, I_s)$, so that $P \equiv 0(A, I_1, \cdots, I_s)$. Since P and Q are relatively prime, then, $y_1^{n-2} \equiv 0(A, I_1, \cdots, I_s)$, so that

$$\left(\frac{P}{y-\eta_i}\right)^3 \equiv 0 \quad (A, I_1, \cdots, I_s, J_1, \cdots, J_s), \qquad i = 1, \cdots. s.$$

From this it follows that $1 \equiv 0 \quad (A, I_1, \cdots, I_s, J_1, \cdots, J_s)$, so that A is of the regular type.

The coefficient in A of y_1^2 is divisible by $y - \eta_i$ if and only if $n > 2$ (in which case the coefficient vanishes). Thus, the singular solutions of A are all of the first class or are all of the second class according as $n > 2$ or $n = 2$. Consequently, the exponent of $Qy_1^n + P$ is 1 or 2 according as $n > 2$ or $n = 2$.

EXAMPLE 4. Let $A = xy_1^2 - 2yy_1 + y$, where x is any element whose derivative is 1. Then $S = 2(xy_1 - y)$, so that the singular solutions of A are $y = 0$ and $y = x$. We may write $A = x(y_1 - 1)^2 - 2(y - x)(y_1 - 1) - (y - x)$. Both singular solutions are of multiplicity unity, and are of the second class.

If we let $\eta_1 = 0$, $\eta_2 = x$, we have, referring to (20) and (18), $I_1 = 2(xy_1 - 2y + x)$, $I_2 = -2(xy_1 - 2y)$. Hence $1 = (I_1 + I_2)/2x$, so that A is of the regular type, and the exponent of $[A]$ is 2.

The following example shows that not every form with singular solutions, all of multiplicity unity, is of the regular type.

EXAMPLE 5. Let $A = y_1^4 + y_1^3 + P$, where P is a form of order zero, with constant coefficients, and with distinct zeros $\eta_1, \cdots, \eta_s$. The singular solutions of A are $\eta_1, \cdots, \eta_s$, all of multiplicity unity and of the first class. $S = 4y_1^3 + 3y_1^2$. By (18) and (20),

$$I_i = \left(\frac{P}{y-\eta_i}\right)^2 (4y_1^2 + 3y_1),$$

$$J_i = \frac{P}{y - \eta_i}\left(-\left(\frac{P}{y - \eta_i}\right)' y_1(y_1^3 + y_1^2) + \left(\frac{P}{y - \eta_i}\right)^2\right)$$

$$\equiv \frac{P}{y - \eta_i}\left(P\left(\frac{P}{y - \eta_i}\right)' + \left(\frac{P}{y - \eta_i}\right)^2\right) \equiv \left(\frac{P}{y - \eta_i}\right)^2 P', \qquad (A).$$

It follows by (29) that $\Sigma_0 = [A, B] = (A, [B])$, where $B = (4y_1^2 + 3y_1)y_2 + P'$. Hence

$$\begin{aligned}(\Sigma_0, S, I_1, \cdots, I_s) &= (y_1^4 + y_1^3 + P, 4y_1^2 + 3y_1, B, B_1, B_2, \cdots) \\ &= (256P^2 - 27P, 9y_1 + 64P, B, B_1, B_2, \cdots) \\ &= (256P - 27, P', 9y_1 + 64P, B_1, B_2, \cdots).\end{aligned}$$

Thus A is of the regular type if $256P - 27$ and P' are relatively prime. However, if $256P - 27$ and P' have a factor in common which is relatively prime to P'', then A is not of the regular type. This is easy to see from the congruences

$$\begin{aligned}B_1 &= (4y_1^2 + 3y_1)y_3 + (8y_1 + 3)y_2^2 + P''y_1 \\ &\equiv -3y_2^2 - \tfrac{3}{4}P'' \qquad (256P - 27, 9y_1 + 64P), \\ B_2 &= (4y_1^2 + 3y_1)y_4 + 3(8y_1 + 3)y_2y_3 + \cdots \\ &\equiv -9y_2y_3 + H_2 \qquad (256P - 27, 9y_1 + 64P), \\ B_k &\equiv -3(k + 1)y_2y_{k+1} + H_k \qquad (256P - 27, 9y_1 + 64P),\end{aligned}$$

where H_k is of order not exceeding k.

11. Case in which no singular solution exists

In this section we suppose that A *has no singular solutions.* By the analog of the Hilbert-Netto theorem this means that

$$1 \equiv 0 \quad (S, S_1, \cdots, S_p, A, A_1, \cdots, A_q)$$

for sufficiently large integers p and q.

We shall consider only those forms A for which p may be taken as 0, that is, for which there is a q such that

$$(45) \qquad 1 \equiv 0 \quad (S, A, A_1, \cdots, A_q).$$

We prove that *if* (45) *holds, than the exponent of* $[A]$ *is unity.*

Indeed, let $G \epsilon \{A\}$. It is required to show that $G \epsilon [A]$. Performing a partial reduction of G with respect to A, we see that there is a positive integer j such that $S^jG \equiv G'[A]$, where G' is of order less than or equal to unity. $G' \epsilon \{A\}$, and hence is divisible by A. Therefore $S^jG \equiv 0$ $[A]$. Of course, $A_iG \equiv 0$ $[A]$, $i = 0, 1, \cdots, q$. By (41), however, $1 \equiv 0$ $(S^j, A, A_1, \cdots, A_q)$. Hence, $G = 1 \cdot G \equiv 0$ $[A]$, q.e.d.

The following example shows that a form A, which has no singular solution, for which (45) fails to hold, is exceptional.

EXAMPLE 6. Let $\mathfrak{F}$ be the differential field obtained by adjoining the unknowns $a_{\mu\nu}$ ($\mu = 0, \cdots, m$; $\nu = 0, \cdots, n$) to some differential field of characteristic zero. We consider in $\mathfrak{F}\{y\}$ the form

$$A = \sum_{\mu=0}^{m} \sum_{\nu=0}^{n} a_{\mu\nu} y^{\mu} y_1^{\nu}.$$

We shall show that A has no singular solutions, and that (45) holds, with $q = 1$.

A_1 is linear in y_2. Let $A_1 = Sy_2 + T$. We must prove that $1 \equiv 0(A, S, T)$. Suppose $1 \not\equiv 0(A, S, T)$. Then there is a solution $y = \bar{y}$, $y_1 = \bar{y}_1$ of the polynomials A, S, T. Now, $\bar{y}$ and $\bar{y}_1$ must annul, respectively, the resultant with respect to y_1 and the resultant with respect to y, of S and T. These resultants are nonzero polynomials in y and y_1, respectively, with coefficients which are polynomials in the $a_{\mu\nu}$ and $a_{\mu\nu 1}$, *not including* a_{00}. It follows, since $y = \bar{y}$, $y_1 = \bar{y}_1$ is a solution of A, that a_{00} is an algebraic function of the other $a_{\mu\nu}$ and the $a_{\mu\nu 1}$. This contradiction completes the proof.

There exist forms A for which (45) is satisfied for some $q > 1$, but for which (45) is not satisfied if $q = 1$. This is shown by

EXAMPLE 7. Let $A = 1 + 2y_1 + (1 - y^3)y_1^2$. Then $S = 2 + 2(1 - y^3)y_1$, and A has no singular solutions. Now, $y_1 + 1 = A - (y_1/2)S$, so that

$$(S, A) = (y^3, y_1 + 1).$$

Also, $A_1 \equiv -3y^2y_1^3 + 2(1 + (1 - y^3)y_1)y_2 \equiv 3y^2(S, A)$. Hence

$$(S, A, A_1) = (y^2, y_1 + 1).$$

Again, $A_2 = -6yy_1^4 - 15y^2y_1^2y_2 + 2(1 - y^3)y_2^2$

$$+ 2(1 + (1 - y^3)y_1)y_3 \equiv -6y + 2y_2^2 \; (S, A, A_1),$$

so that

$$(S, A, A_1, A_2) = (y^2, y_1 + 1, y_2^2 - 3y).$$

Finally, $A_3 = -6y_1^5 - 54yy_1^3y_2 - 36y^2y_1y_2^2 - 21y^2y_1^2y_3 + 6(1 - y^3)y_2y_3 + 2(1 + (1 - y^3)y_1)y_4 \equiv 6(1 + 9yy_2 - y_2y_3)$, (S, A, A_1, A_2), so that $1 \equiv y_2(y_3 - 9y)$, (S, A, A_1, A_2, A_3), and $1 \equiv y_2^4(y_3 - 9y)^4 \equiv 9y^2(y_3 - 9y)^4 \equiv 0 \; (S, A, A_1, A_2, A_3)$. Thus, (45) holds for $q = 3$, but for no lower value of q.

12. The intermediate ideals for hyperelliptic forms

We have seen in §10 that if A has singular solutions, all of multiplicity unity, at least one of which is of the second class, and if A is of the regular type, then $[A]$ is properly included in $\{A\}$, and has exponent 2. It is natural now to ask what differential ideals there are *between* $[A]$ and $\{A\}$, that is, what differential ideals Λ there are such that $[A] \subseteq \Lambda \subseteq \{A\}$. We shall call such a Λ an *intermediate ideal.* In the present section we answer this question completely for the forms of Example 3 with $n = 2$.

We are dealing then with a form $A = Qy_1^2 + P$, where P and Q have constant

coefficients, are of order zero, and are relatively prime. P is assumed to be of positive degree, relatively prime to $P' = \frac{d}{dy}P$. Such a form A we shall call a *hyperelliptic* form.

Our first task will be to find a suitable basis for Σ_0.

Let $\eta_1, \cdots, \eta_s$ be the singular solutions of A, that is, the roots of $P = 0$. We have seen that

$$(46) \qquad \Sigma_0 = [A, (I_1y_2 + J_1)^2, \cdots, (I_sy_2 + J_s)^2],^{32}$$

where

$$(47) \qquad (y - \eta_i)(I_iy_2 + J_i) \equiv 0, \qquad y_1(I_iy_2 + J_i) \equiv 0 \qquad [A],$$

and

$$(48) \qquad 1 \equiv 0\ (A, I_1, \cdots, I_s).$$

Congruence (48) shows that there is a form K, with constant coefficients and of order less than 2, such that

$$(49) \qquad y_2 + K \equiv 0\ (A, I_1y_2 + J_1, \cdots, I_sy_2 + J_s).$$

We shall show that

$$(50) \qquad \Sigma_0 = [A, (y_2 + K)^2], \qquad \{\Sigma_0\} = [A, y_2 + K],$$

and that

$$(51) \qquad (y - \eta_1) \cdots (y - \eta_s)(y_2 + K) \equiv 0, \qquad y_1(y_2 + K) \equiv 0 \qquad [A].$$

Now, (51) is obvious from (47) and (49). As to (50), it is evident from (46), (49) and the fact that Σ_0 is of exponent 2, that $[A, (y_2 + K)^2] \subseteq \Sigma_0$. To prove the inclusion in the opposite direction, consider any $I_iy_2 + J_i$. We have

$$I_iy_2 + J_i = I_i(y_2 + K) + J_i - I_iK.$$

Hence $J_i - I_iK \in \{\Sigma_0\}$. Since the order of $J_i - I_iK$ does not exceed unity, this form is divisible by A. Thus,

$$I_i\, y_2 + J_i \equiv 0\ (A, y_2 + K),$$

so that

$$(I_iy_2 + J_i)^2 \equiv 0\ [A, (y_2 + K)^2], \qquad i = 1, \cdots, s.$$

This, in conjunction with (46), shows that $\Sigma_0 \subseteq [A, (y_2 + K)^2]$, and completes the proof of the first part of (50). The second part of (50) is now obvious.

The congruence (50) gives the basis sought for Σ_0. We have, as a result,

[32] Since P has constant coefficients the η_i are constants, so that $(y - \eta_i)_j = y_j$, $i = 1, \cdots, s$, $j = 1, 2, \cdots$.

(52) $$[A] = [P]\cdot[A, (y_2 + K)^2], \qquad \{A\} = [P]\cdot[A, y_2 + K],$$

for P is a constant times the product of the $y - \eta_i$.[33]

If we refer to (51), the first intermediate ideals which present themselves are of the type

(53) $$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_g})(y_2 + K)],$$

where $i_1, \cdots, i_g$ are distinct integers from 1 to s. If $g = s$, the ideal (53) is $[A]$, if $g = 0$, (53) is $\{A\}$.

We shall show that *the ideals* (53) *are distinct from one another*, and that *every intermediate ideal is one of the ideals* (53), that is, the ideals (53) are the only intermediate ideals.

To prove that the ideals (53) are all distinct, it suffices to show, $h = 1, \cdots, s$, that

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{h-1}})(y_2 + K)]$$
$$\neq [P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_h})(y_2 + K)].$$

For, if

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_h})(y_2 + K)]$$

and

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{j_1}) \cdots (y - \eta_{j_f})(y_2 + K)]$$

were equal, they would both equal

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{k_1}) \cdots (y - \eta_{k_l})(y_2 + K)],$$

where $k_1, \cdots, k_l$ are the integers common to $i_1, \cdots, i_h$ and $j_1, \cdots, j_f$; and if

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{h-d}})(y_2 + K)]$$
$$= [P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_h})(y_2 + K)],$$

with $d > 1$, then we would have

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{h-1}})(y_2 + K)$$
$$= [P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_h})(y_2 + K)].$$

For $h = s$ this means we must show that

$$[A] \subset [P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_2 + K)]$$

for all distinct $i_1, \cdots, i_{s-1}$. This will be proved, however, if we show that

(54) $$(y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_3 + K_1) \not\equiv 0 \qquad [A],$$[34]

for the first member here is in

[33] See (10).

[34] K_j is the j-th derivative of K.

$$[P]\cdot[A,\ (y_2+K)^2,\ (y-\eta_{i_1})\cdots(y-\eta_{i_{s-1}})(y_2+K)].^{35}$$

The relation (54) may be established in the same way in which (37) was disproved.

From (54) it easily follows that

$$(y-\eta_{i_1})\cdots(y-\eta_{i_{s-1}})(y_2+K)\not\equiv 0 \qquad (\Sigma_0).$$

For $h < s$ we proceed by induction. Assume that

$$\begin{aligned}(55)\qquad (y-\eta_{i_1})\cdots(y-\eta_{i_{j-1}})(y-\eta_{i_{j+1}})\cdots(y-\eta_{i_{h+1}})(y_2+K)&\\ \not\equiv 0\ \ [A,\ (y_2+K)^2,\ (y-\eta_{i_1})\cdots(y-\eta_{i_{h+1}})(y_2+K)]&\end{aligned}$$

for all distinct $i, \cdots, i_{h+1}$ and all $j = 1, \cdots, h+1$. (For $h = s-1$ this has just been proved.)

Suppose, say, that

$$\begin{aligned}(y-\eta_1)\cdots(y-\eta_{h-1})(y_2+K)&\\ \equiv 0\ \ [A,\ (y_2+K)^2,\ (y-\eta_1)\cdots(y-\eta_h)(y_2+K)].&\end{aligned}$$

Then, for t sufficiently large,

$$\begin{aligned}(y-\eta_1)\cdots(y-\eta_{h-1})(y-\eta_{h+1})^t(y_2+K)&\\ \equiv 0\ \ [A,\ (y_2+K)^2,\ (y-\eta_1)\cdots(y-\eta_{h+1})(y_2+K)].&\end{aligned}$$

Combining this congruence with

$$\begin{aligned}(y-\eta_1)\cdots(y-\eta_{h+1})(y_2+K)&\\ \equiv 0\ \ [A,\ (y_2+K)^2,\ (y-\eta_1)\cdots(y-\eta_{h+1})(y_2+K)],&\end{aligned}$$

we see that

$$\begin{aligned}(y-\eta_1)\cdots(y-\eta_{h-1})(y-\eta_{h+1})(y_2+K)&\\ \equiv 0\ \ [A,\ (y_2+K)^2,\ (y-\eta_1)\cdots(y-\eta_{h+1})(y_2+K)].&\end{aligned}$$

This contradicts (55), and completes the first part of the proof.

As for the second part, let Σ be a differential ideal between Σ_0 and $\{\Sigma_0\}$, distinct from $\{\Sigma_0\}$. We shall complete the proof by showing that Σ coincides with some ideal

$$(56)\qquad [A,\ (y_2+K)^2,\ (y-\eta_{i_1})\cdots(y-\eta_{i_q})(y_2+K)].$$

[35] The first member of (54) is in $[P]$ because it is in $[y_1]$ and $y_1 \epsilon [P]$; the first member of (54) is in $[A,\ (y^2+K)^2,\ (y-\eta_{i_1})\cdots(y-\eta_{i_{s-1}})(y_2+K)]$ because $((y-\eta_{i_1})\cdots(y-\eta_{i_{s-1}})(y_2+K))_1 = ((y-\eta_{i_1})\cdots(y-\eta_{i_{s-1}}))'y_1(y_2+K) + (y-\eta_{i_1})\cdots(y-\eta_{i_{s-1}})(y_2+K_1) \equiv (y-\eta_{i_1})\cdots(y-\eta_{i_{s-1}})(y_2+K_1)$, (Σ_0). Moreover, the ideals $[P]$ and $[A,\ (y_2+K)^2,\ (y-\eta_{i_1})\cdots(y-\eta_{i_{s-1}})(y_2+K)]$ are separated, because P and y_2+K have no solution in common.

Let q be the smallest integer for which some ideal (56) with $g = q$ is included in Σ. q is an integer from 1 to s. Suppose, say, that

$$\Lambda \subseteq \Sigma,$$

where

$$\Lambda = [A, (y_2 + K)^2, (y - \eta_1) \cdots (y - \eta_q)(y_2 + K)],$$

but that

$$[A, (y_2 + K)^2, (y - \eta_1) \cdots (y - \eta_{i-1})(y - \eta_{i+1}) \cdots (y - \eta_q)(y_2 + K)] \nsubseteq \Sigma,$$

$i = 1, \cdots, q$. We shall prove that $\Sigma = \Lambda$.

Assume to the contrary that $F \in \Sigma$, $F \notin \Lambda$. Since $F \in \{\Sigma_0\} = [A, y_2 + K]$, we may write

$$F = C_0 A + C_1(y_2 + K) + \cdots + C_t(y_{t+1} + K_{t-1}). \tag{57}$$

Let D_t be the remainder of C_t with respect to $y_2 + K$. Since $(y_{i+1} + K_{i-1})(y_{t+1} + K_{t-1}) \equiv 0 \ (\Lambda)$, $i = 1, 2, \cdots$, we have, by (57),

$$F \equiv C_1(y_2 + K) + \cdots + C_{t-1}(y_t + K_{t-2}) + D_t(y_{t+1} + K_{t-1}), \qquad (\Lambda). \tag{58}$$

D_t is a polynomial in y, y_1. Let E_t be the sum of the terms of D_t free of y_1. Since, by (51), $(y_1(y_2 + K))_{t-1} \equiv 0 \ (\Lambda)$, it follows that

$$y_1(y_{t+1} + K_{t-1}) \equiv -(t-1)y_2(y_t + K_{t-2}) - \cdots - y_t(y_2 + K), \qquad (\Lambda). \tag{59}$$

From this and (58) we see that

$$F \equiv C_1'(y_2 + K) + \cdots + C_{t-1}'(y_t + K_{t-2}) + E_t(y_{t-1} + K_{t-1}), \qquad (\Lambda), \tag{60}$$

where the C_i' are new forms.

Let H_t be the remainder obtained on dividing E_t by $(y - \eta_1) \cdots (y - \eta_q)$. Since $((y - \eta_1) \cdots (y - \eta_q)(y_2 + K))_{t-1} \equiv 0 \ (\Lambda)$,

$$\begin{aligned}(y - \eta_1) \cdots (y - \eta_q)(y_{t+1} + K_{t-1}) \\ \equiv -(t-1)((y - \eta_1) \cdots (y - \eta_q))_1(y_t + K_{t-1}) \\ - \cdots - ((y - \eta_1) \cdots (y - \eta_q))_{t-1}(y_2 + K), \qquad (\Lambda).\end{aligned}$$

Hence, it follows from (60) that

$$F \equiv C_1''(y_2 + K) + \cdots + C_{t-1}''(y_t + K_{t-2}) + H_t(y_{t+1} + K_{t-1}), \qquad (\Lambda), \tag{61}$$

where the C_i'' are new forms.

Treating the first $t - 1$ terms in the second member of (61) as we did the second member of (57), we obtain, after $t - 1$ more steps, a relation

$$F \equiv H_1(y_2 + K) + \cdots + H_t(y_{t+1} + K_{t-1}), \qquad (\Lambda), \tag{62}$$

where each H_i is a form of order zero and degree less than q.

At least one H_i must be nonzero, else $F \in \Lambda$. We may (and do) suppose that $H_t \neq 0$.

From the congruence $y_1(y_2 + K) \equiv 0 \quad (\Lambda)$, we obtain relations

$$y_1^j(y_{j+1} + K_{j-1}) \equiv 0 \qquad (\Lambda), \qquad j = 1, 2, \cdots. \tag{63}$$

Therefore, if we multiply (62) by y_1^{t-1} we find

$$H_t y_1^{t-1}(y_{t+1} + K_{t-1}) \equiv 0 \qquad (\Sigma).$$

Now, by (59) and (63),

$$\begin{aligned} y_1^{j-1}(y_{j+1} + K_{j-1}) &\equiv -(j-1)y_1^{j-2}y_2(y_j + K_{j-2}) \\ &\equiv (j-1)y_1^{j-2}K(y_j + K_{j-2}), \qquad (\Lambda), \end{aligned}$$

so that

$$y_1^{t-1}(y_{t+1} + K_{t-1}) \equiv (t-1)!K^{t-1}(y_2 + K), \qquad (\Lambda).$$

Hence,

$$H_t K^{t-1}(y_2 + K) \equiv 0 \qquad (\Sigma).$$

If L is the sum of the terms of K free of y_1, we find, since $y_1(y_2 + K) \equiv 0$ (Σ)

$$H_t L^{t-1}(y_2 + K) \equiv 0 \qquad (\Sigma).$$

But $(y - \eta_1) \cdots (y - \eta_q)(y_2 + K) \equiv 0$ (Σ). Hence, if J is the greatest common divisor of $H_t L^{t-1}$ and $(y - \eta_1) \cdots (y - \eta_q)$,

$$J(y_2 + K) \equiv 0 \qquad (\Sigma).$$

But L is prime to $(y - \eta_1) \cdots (y - \eta_q)$, for otherwise the ideals $[y - \eta_1], \cdots, [y - \eta_q], [A, y_2 + K]$ would not be separated. Hence J is a divisor of H_t, is of degree less than q, and is a *proper* divisor of $(y - \eta_1) \cdots (y - \eta_q)$. This contradicts the definition of q as the least g for which Σ includes an ideal (56), and completes the proof.

The results just established show that the ideals between $[A]$ and $\{A\}$ can be organized into ascending *chains*, all of the same length s. That is, if $\Lambda_1, \cdots, \Lambda_{h-1}$ are intermediate ideals such that

$$[A] \subset \Lambda_1 \subset \cdots \subset \Lambda_{h-1} \subset \{A\}$$

(chain of length h), then there are $s - h$ other intermediate ideals which together with $[A], \Lambda_1, \cdots, \Lambda_{h-1}, \{A\}$ form a chain of length s. Moreover, there exist no chains of length exceeding s. This illustrates a phenomenon which we shall discuss in Part III.

Part III. Chains

13. Generalities

We return now to an abstract differential domain of integrity with basis theorem, which contains the rational numbers.

The $n + 1$ differential ideals $\sigma_0, \sigma_1, \cdots, \sigma_{n-1}, \sigma^*$ are said to form a *chain* if

$$\sigma_0 \subset \sigma_1 \subset \cdots \subset \sigma_{n-1} \subset \sigma^*. \tag{64}$$

n is called the *length* of the chain. We shall call σ_0 and σ^* the *end ideals*, $\sigma_1, \cdots, \sigma_{n-1}$ the *intermediate ideals*, of the chain. The ideals σ_{i-1}, σ_i (also, σ_{n-1}, σ^*) will be said to be *contiguous*. A set of ideals of the chain will be called *non-contiguous* if no two ideals of the set are contiguous. A chain is called a *principal* chain if no pair of contiguous ideals have a differential ideal properly between them. A principal chain may be expressed schematically by means of dashes:

$$\sigma_0 - \sigma_1 - \cdots - \sigma_{n-1} - \sigma^*.$$

The notation $\sigma - \tau$ means that $\sigma \subset \tau$, but there is no ω with $\sigma \subset \omega \subset \tau$. A chain with end ideals σ_0, σ^* is said to be a *refinement* of the chain (64) if each intermediate ideal of (64) is an intermediate ideal of the chain in question.

Let $\sigma_0, \sigma_1, \cdots, \sigma_{n-1}, \sigma^*$ be a principal chain, and let there be a second principal chain $\sigma_0, \sigma_1, \cdots, \sigma_{i-1}, \tau_i, \sigma_{i+1}, \cdots, \sigma_{n-1}, \sigma^*$ with $\sigma_i \neq \tau_i$, that is, schematically, let

$$\sigma_0 - \sigma_1 - \cdots - \sigma_{i-1} \begin{array}{c} \nearrow \sigma_i \searrow \\ \searrow \tau_i \nearrow \end{array} \sigma_{i+1} - \cdots - \sigma_{n-1} - \sigma^*. \tag{65}$$

We shall say that the first chain is *flexible* at the intermediate ideal σ_i, and that the second chain is obtained from the first by a *flex*.

The first theorem on chains that we shall state is that of Jordan-Hölder-Schreier. A differential domain of integrity $\mathfrak{D}$ forms a group under addition. As a set of operators on this group we may consider 1. multiplication by an element of $\mathfrak{D}$, 2. differentiation. The permissible invariant subgroups of this group with operators are precisely the differential ideals of $\mathfrak{D}$. Thus, we may state

THEOREM 7 (Jordan-Hölder-Schreier): *Two chains with the same end ideals have refinements of equal length, with differential rings of remainder classes which are isomorphic (in some order).*[36]

PROOF: See van der Waerden, *Moderne Algebra*, vol. 1, 2nd edition, §46, or H. Zassenhauss, *Lehrbuch der Gruppentheorie*, vol. 1, chapter 2, §5.

As a consequence of Theorem 7, we see that all principal chains with the same end ideals have the same length.

THEOREM 8 (Ore): *Two different principal chains with the same end ideals, can be obtained from each other by a finite number of successive flexes.*

PROOF: See O. Ore, *On the foundations of abstract algebra.* I, Annals of Mathematics, vol. 36 (1935), pp. 406–434.

14. The exponent of a principal chain

By the *exponent* of the chain (64) we shall mean $l_{\sigma^*}\sigma_0$. We shall be concerned with principal chains, exclusively. Of course, if $\sigma^* \not\subseteq \{\sigma_0\}$, then the chain (64) has exponent ∞, so we may limit ourselves to chains for which $\sigma^* \subseteq \{\sigma_0\}$.

---

[36] By the differential ring of remainder classes of a chain (64) is meant the differential rings $\sigma_1/\sigma_0, \sigma_2/\sigma_1, \cdots, \sigma^*/\sigma_{n-1}$.

Our chief result may be stated as follows.

THEOREM 9: Let

$$(66)\qquad \sigma_0 - \sigma_1 - \cdots - \sigma_{n-1} - \sigma^*, \qquad \sigma^* \subseteq \{\sigma_0\},$$

be a principal chain which is flexible at r non-contiguous intermediate ideals.[37] *Then the exponent of the chain* (66) *does not exceed* $n - r + 1$.

As a consequence of Theorem 9, we see that if a differential ideal σ has finite exponent p, then there is a chain, with end ideals σ and $\{\sigma\}$, of length $p - 1$.

PROOF: We begin by showing that, if $\sigma_0 - \sigma_1$, and $\sigma_1 \subseteq \{\sigma_0\}$, then $l_{\sigma_1}\sigma_0 = 2$. Indeed, let $l_{\sigma_1}\sigma_0 > 2$. Then there is a $b \in \sigma_1$ such that $b^2 \notin \sigma_0$. Hence

$$\sigma_0 \subset [\sigma_0, b^2] \subseteq \sigma_1 .$$

Since $\sigma_0 - \sigma_1$, we see that $\sigma_1 = [\sigma_0, b^2]$. In particular, $b \in [\sigma_0, b^2]$, so that

$$b \equiv A\ (\sigma_0),$$

where A is a form in b with coefficients in $\mathfrak{D}$, which has no term of degree less than 2. Using the iterative process introduced in the paper cited under footnote 19, we see that $b \equiv 0\ (\sigma_0)$. This contradiction shows that $l_{\sigma_1}\sigma_0 = 2$.

We now show that the exponent of the principal chain (66) does not exceed $n + 1$.

For chains of length 1 we have just given the proof. Let $n > 1$, and assume the result for chains of length less than n.

By the assumption, $\sigma^{*n} \subseteq \sigma_1$. Suppose $\sigma^{*n+1} \nsubseteq \sigma_0$. Then

$$\sigma_0 \subset (\sigma_0, \sigma^{*n+1}) \subseteq \sigma_1 .$$

Since $\sigma_0 - \sigma_1$, we see that $\sigma_1 = (\sigma_0, \sigma^{*n+1})$. Hence

$$\begin{aligned} \sigma^{*n+1} &= \sigma^*\sigma^{*n} \subseteq \sigma^*\sigma_1 = \sigma^*(\sigma_0, \sigma^{*n+1}) \\ &\subseteq (\sigma_0, \sigma^{*n+2}) \subseteq (\sigma_0, \sigma^{*2}\sigma_1) \\ &\subseteq (\sigma_0, (\sigma_0, \sigma^{*n+3})) = (\sigma_0, \sigma^{*n+3}) \\ &\subseteq \cdots \subseteq (\sigma_0, \sigma^{*2n}) \subseteq (\sigma_0, \sigma_1^2) \subseteq \sigma_0 . \end{aligned}$$

This contradiction proves our statement.

We are now ready to complete the proof of our theorem. What we have just shown is that the theorem is valid for $r = 0$. We proceed by induction. Let $r > 0$ and suppose the theorem holds for fewer than r non-contiguous ideals.

Let σ_i be the intermediate ideal of (66) of lowest subscript, at which the chain is flexible. Then we have, with $\sigma_i \neq \tau_i$, the situation depicted by (65). Now

$$\sigma_i - \sigma_{i+1} - \cdots - \sigma_{n-1} - \sigma^*$$

[37] Clearly, r is an integer between 0 and the greatest integer in $n/2$, inclusive.

and

$$\tau_i - \sigma_{i+1} - \cdots - \sigma_{n-1} - \sigma^*$$

are both principal chains of length $n - i$, flexible at $r - 1$ intermediate ideals. Hence, by the induction assumption, $\sigma^{*^{n-i-r+2}} \subseteq \sigma_i$, $\sigma^{*^{n-i-r+2}} \subseteq \tau_i$. But $\sigma_{i-1} \subseteq \sigma_i \cap \tau_i \subset \sigma_i$ and $\sigma_{i-1} - \sigma_i$, so that $\sigma_{i-1} = \sigma_i \cap \tau_i$. Therefore, $\sigma^{*^{n-i-r+2}} \subseteq \sigma_{i-1}$.

Suppose, now, that $\sigma^{*^{n-r+1}} \not\subseteq \sigma_0$. Let σ_k be the σ_j of least subscript, such that $\sigma^{*^{n-j-r+1}} \subseteq \sigma_j$. Then $1 \leqq k \leqq i - 1$ and

$$\sigma^{*^{n-k-r+1}} \subseteq \sigma_k, \qquad \sigma^{*^{n-k-r+2}} \not\subseteq \sigma_{k-1}, \qquad \sigma^{*^{n-k-r+3}} \subseteq \sigma_{k-1}.$$

Hence

$$\sigma_{k-1} \subset (\sigma_{k-1}, \sigma^{*^{n-k-r+2}}) \subseteq \sigma_k,$$

so that, since $\sigma_{k-1} - \sigma_k$, $\sigma_k = (\sigma_{k-1}, \sigma^{*^{n-k-r+2}})$. Thus,

$$\begin{aligned} \sigma^{*^{n-k-r+2}} &= \sigma^* \sigma^{*^{n-k-r+1}} \subseteq \sigma^* \sigma_k \\ &= \sigma^*(\sigma_{k-1}, \sigma^{*^{n-k-r+2}}) \\ &\subseteq (\sigma_{k-1}, \sigma^{*^{n-k-r+3}}) = \sigma_{k-1}. \end{aligned}$$

This contradiction completes the proof of Theorem 9.

EXAMPLE 8. Consider the chain

$$[y^{n+1}, y_1], [y^n, y_1], \cdots, [y^2, y_1], [y]$$

in any differential ring $\mathfrak{F}\{y\}$. It is easy to prove that this is a principal chain, flexible at none of its intermediate ideals. The length is n and the exponent is obviously $n + 1$. Thus, the bound on the exponent given by Theorem 9 is the exponent itself.

The necessity of the non-contiguity condition in Theorem 9 is shown by

EXAMPLE 9. Let $\mathfrak{F}$ be a differential field of characteristic zero, which contains an element x with derivative 1. Consider the chain

$$[y^3, yy_1, y_2], [y^2, y_1^2, y_2], [y^2, y_1], [y]$$

in $\mathfrak{F}\{y\}$. This is a principal chain of length 3. Since

$$\begin{tikzcd}[column sep=small, row sep=small] & {[y^2, y_1^2, y_2]} \arrow[dr, no head] & \\ {[y^3, yy_1, y_2]} \arrow[ur, no head] \arrow[dr, no head] & & {[y^2, y_1]} \\ & {[y^3, y_1]} \arrow[ur, no head] & \end{tikzcd}$$

and

$$\begin{tikzcd}[column sep=small, row sep=small] & {[y^2, y_1]} \arrow[dr, no head] & \\ {[y^2, y_1^2, y_2]} \arrow[ur, no head] \arrow[dr, no head] & & {[y],} \\ & \left[y^2, y_1 - \frac{1}{x} y\right] \arrow[ur, no head] & \end{tikzcd}$$

the chain is flexible at two intermediate ideals. Yet its exponent is 3.

Appendix. General Solutions and Resolvents

Liquidating an obligation incurred in Part I, we give here an abstract, purely algebraic treatment of the subject matter of Chapter 2 of Ritt's book, *Differential equations from the algebraic standpoint.*[38] We shall make free reference to this book, designating it by (*R*), and shall give proofs only when they differ from those in (*R*).

1*. The general solution of a form

We work in an abstract differential field $\mathfrak{F}$ of characteristic zero. The letters u_i, y_i and w will denote unknowns.

We establish the following fundamental lemma as a substitute for the existence theorem for differential equations.

Lemma 1*: Let A be an algebraically irreducible form in $\mathfrak{F}\{y_1, \cdots, y_n\}$ of positive class. Let $B_1, \cdots, B_t$ be nonzero forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$, reduced with respect to A. Then $B_1 \cdots B_t \notin \{A\}$.

Remark: Lemma 1*, together with the analog to the Hilbert-Netto theorem, shows that every algebraically irreducible form of positive class has a *regular* solution, that is, a solution for which the separant S and initial I of A do not vanish. For otherwise IS would be in $\{A\}$.

Proof: Let m be the class of A, r the order of A in y_m. Suppose $B_1 \cdots B_t \in \{A\}$, that is, $(B_1 \cdots B_t)^p \in [A]$, for some positive integer p. Now, A and $(B_1 \cdots B_t)^p$ are relatively prime. Hence there is a nonzero form C, either of class less than m, or of class m and of order less than r in y_m, such that

$$C = DA + E(B_1 \cdots B_t)^p,$$

where D and E are forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$. $C \in [A]$, so that we may write, for suitable q and Q_i,

$$C = Q_0A + Q_1A_1 + \cdots + Q_qA_q. \tag{1*}$$

Now, $A_i = Sy_{m,r+i} + T_i$, where T_i is a form of order less than $r + i$ in y_m. Hence, if in (1*) we consider the forms as polynomials in the letters y_{ij}, and let

$$y_{m,r+i} = -\frac{T_i}{S}, \qquad i = 1, 2, \cdots,$$

we see that S^hC is divisible by A for a sufficiently large integer h. Since neither S nor C is so divisible, we have a contradiction, completing the proof.

We now are in a position to follow (*R*).

Let $A \in \mathfrak{F}\{y_1, \cdots, y_n\}$ be algebraically irreducible and of class n. Let S and I be the separant and initial, respectively, of A. Let Σ_1 be the set of all forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$ which vanish for all regular solutions of A, or what is the same thing, let $\Sigma_1 = \{A\}:IS$.

[38] American Mathematical Society Colloquium Publications, vol. 14.

Following (R), we can show that Σ_1 *is a prime differential ideal.* The manifold of Σ_1 is called the *general solution* of A.

The following three results are proved as in (R):

$\Sigma_1 = \{A\}:S$;

Σ_1 *is an essential prime differential ideal in the decomposition of* $\{A\}$; *the other essential ideals contain* S;

Σ_1 *is independent of the order of the unknowns.*

2*. The resolvent of a prime differential ideal

Let Σ be a perfect differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Let

$$A_1, \cdots, A_r \tag{2*}$$

be a basic set of Σ. We denote the separant and initial of A_i, $i = 1, \cdots, r$, by S_i and I_i, respectively. Let $\Sigma' = \{A_1, \cdots, A_r\}:I_1S_1 \cdots I_rS_r$, that is, let Σ' be the set of all forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$ which vanish for all regular solutions of the basic set (2*).[39] Clearly, $\Sigma \subseteq \Sigma'$.

Following (R), we see that *if* Σ *is prime, then* $\Sigma = \Sigma'$. In particular (2*) has regular solutions. Also, Σ *is the only prime differential ideal with basic set* (2*).

Let $\mathfrak{C}$ be the set of all constants (that is, elements with vanishing derivative) in $\mathfrak{F}$. $\mathfrak{C}$ itself is a differential field of characteristic zero.

We shall need the following

LEMMA 2*: $\eta_1, \cdots, \eta_n \in \mathfrak{F}$ *are linearly dependent over* $\mathfrak{C}$ *if and only if*

$$\begin{vmatrix} \eta_1 & \cdots & \eta_n \\ \eta_{11} & \cdots & \eta_{n1} \\ \cdots & \cdots & \cdots \\ \eta_{1,n-1} & \cdots & \eta_{n,n-1} \end{vmatrix} = 0. \tag{3*}$$

PROOF: If $\sum_{\nu=1}^{n} c_\nu \eta_\nu = 0$, where the c_ν are in $\mathfrak{C}$ and not all 0, then $\sum_{\nu=1}^{n} c_\nu \eta_{\nu i} = 0$, $i = 0, 1, \cdots, n - 1$, so that (3*) must hold. Conversely, let (3*) hold. If $n = 1$, the lemma is obviously valid. Let $n > 1$ and suppose the lemma holds for fewer than n elements η_ν. Let

$$\mu_\nu = (-1)^{n+\nu} \begin{vmatrix} \eta_1 & \cdots & \eta_{\nu-1} & \eta_{\nu+1} & \cdots & \eta_n \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \eta_{1,n-2} & \cdots & \eta_{\nu-1,n-2} & \eta_{\nu+1,n-2} & \cdots & \eta_{n,n-2} \end{vmatrix}, \qquad \nu = 1, \cdots, n.$$

If $\mu_n = 0$, the lemma for $n - 1$ shows that the η_ν are linearly dependent over $\mathfrak{C}$. Assume that $\mu_n \neq 0$. By (3*),

$$\sum_{\nu=1}^{n} \eta_{\nu i} \mu_\nu = 0, \qquad i = 0, 1, \cdots, n - 1.\text{[40]} \tag{4*}$$

[39] A solution of (2*) is called *regular* if it does not annul $I_1S_1 \cdots I_rS_r$.

[40] For $i = n - 1$, the first member of (4*) is the development of the determinant in (3*) by the minors of the elements of the last row. For $i < n - 1$, (4*) is a statement of the vanishing of certain determinants with two identical rows.

Differentiating the first $n - 1$ equations (4*), we find

$$\sum_{\nu=1}^{n} \eta_{\nu,i+1}\mu_\nu + \sum_{\nu=1}^{n} \eta_{\nu i}\mu_{\nu 1} = 0, \qquad i = 0, 1, \cdots, n-2.$$

Taking into account the last $n - 1$ equations (4*), we see that $\sum_{\nu=1}^{n} \eta_{\nu i}\mu_{\nu 1} = 0$, $i = 0, 1, \cdots, n - 2$. These equations may be written

$$\sum_{\nu=1}^{n-1} \eta_{\nu i}\mu_{\nu 1} = -\eta_{ni}\mu_{n1}, \qquad i = 0, 1, \cdots, n-2.$$

Solving these equations in $\mu_{11}, \cdots, \mu_{n-1,1}$ by Cramer's rule, we obtain

$$\mu_{\nu 1} = \frac{\mu_{n1}}{\mu_n}\mu_\nu, \qquad \nu = 1, \cdots, n-1.$$

Hence $(\mu_\nu/\mu_n)_1 = (\mu_n\mu_{\nu 1} - \mu_{n1}\mu_\nu)/\mu_n^2 = 0$, so that $\mu_\nu/\mu_n = c_\nu$, $\nu = 1, \cdots, n$, where the c_ν are constants, not all zero ($c_n = 1$). Using (4*) with $i = 0$, we see that $\sum_{\nu=1}^{n} c_\nu\eta_\nu = 0$, which completes the proof.

LEMMA 3*: *Let $\mathfrak{F}$ contain a non-constant element ξ. Let A be a nonzero form in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then there are elements $\eta_1, \cdots, \eta_n \in \mathfrak{F}$ such that $A \neq 0$ when $y_1 = \eta_1, \cdots, y_n = \eta_n$.*

PROOF: It suffices to consider the case $n = 1$, for the cases $n > 1$ follow by induction. Let $A \in \mathfrak{F}\{y\}$ be of order p, and let $Y_j = \sum_{i=0}^{p} c_i(\xi^i)_j$, $j = 0, 1, \cdots, p$, where the c_i are new unknowns.[41] The Jacobian of the $p + 1$ functions Y_j of the $p + 1$ variables c_i is

$$\omega = \begin{vmatrix} 1 & \xi & \xi^2 & \cdots & \xi^p \\ 0 & \xi_1 & (\xi^2)_1 & \cdots & (\xi^p)_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \xi_p & (\xi^2)_p & \cdots & (\xi^p)_p \end{vmatrix}.$$

If ω were 0, ξ would be algebraic over $\mathfrak{C}$ (by Lemma 2*) and would therefore be a constant, so that $\omega \neq 0$. Hence $Y_0, \cdots, Y_p$, considered as polynomials in $c_0, \cdots, c_p$, are algebraically independent.[42] Therefore, if we substitute $Y_0, Y_1, \cdots, Y_p$ in A for $y, y_1, \cdots, y_p$, respectively, A will go over into a nonzero polynomial C in the c_i. Choose integers $\gamma_0, \gamma_1, \cdots, \gamma_p$ such that $C \neq 0$ when $c_i = \gamma_i$, $i = 0, 1, \cdots, p$. Then the element $\eta = \sum_{i=0}^{p} \gamma_i\xi^i \in \mathfrak{F}$ leaves A nonzero when substituted for y.

Using Lemma 3*, we can follow the proof of §25 in (R) to obtain the following result.

Let Σ be a non-trivial prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$ with arbitrary unknowns $u_1, \cdots, u_q$ (q may be 0.) If $\mathfrak{F}$ contains a non-constant, then there are elements $\mu_1, \cdots, \mu_p \in \mathfrak{F}$, and a nonzero form $G \in \mathfrak{F}\{u_1, \cdots, u_q\}$, such that if

[41] $(\xi^i)_j$ is the j-th derivative of ξ^i.

[42] See O. Perron, *Algebra*, vol. 1, Berlin, 1932, p. 134.

(5*) $\bar{u}_1, \cdots, \bar{u}_q, y_1', \cdots, y_p'$ *and* $\bar{u}_1, \cdots, \bar{u}_q, y_1'', \cdots, y_p''$

are two distinct solutions of Σ, *then either* $G(\bar{u}_1, \cdots, \bar{u}_q) = 0$, *or*

$$\mu_1(y_1' - y_1'') + \cdots + \mu_p(y_p' - y_p'') \neq 0.$$

Taking over the proof in §26 of (R), we obtain:

Let $\mathfrak{F}$ *be any differential field of characteristic zero. Let* Σ *be a nontrivial prime differential ideal in* $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$ *with arbitrary unknowns* $u_1, \cdots, u_q$. *We suppose that* $q > 0$. *Then there are forms*

$$G, M_1, \cdots, M_p \in \mathfrak{F}\{u_1, \cdots, u_p\},$$

with $G \neq 0$, *such that if* (5*) *are two distinct solutions of* Σ, *then either* $G(\bar{u}_1, \cdots, \bar{u}_q) = 0$ *or*

$$M_1(\bar{u}_1, \cdots, \bar{u}_q)\cdot(y_1' - y_1'') + \cdots + M_p(\bar{u}_1, \cdots, \bar{u}_q)\cdot(y_p' - y_p'') \neq 0.$$

Henceforth we shall deal with a non-trivial prime differential ideal Σ in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, with arbitrary unknowns $u_1, \cdots, u_q$, and we shall assume that either $\mathfrak{F}$ contains non-constant elements, or the arbitrary unknowns u_i really exist. Then there exists a triad of forms G, P, $Q \in \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, with G free of the y_i, and not 0, and with $P \notin \Sigma$, such that if (5*) are two distinct solutions of Σ for which $GP \neq 0$, then

$$\frac{Q(\bar{u}_1, \cdots, \bar{u}_q, y_1', \cdots, y_p')}{P(\bar{u}_1, \cdots, \bar{u}_q, y_1', \cdots, y_p')} \neq \frac{Q(\bar{u}_1, \cdots, \bar{u}_q, y_1'', \cdots, y_p'')}{P(\bar{u}_1, \cdots, \bar{u}_q, y_1'', \cdots, y_p'')}.$$

Following (R), we form the ideal

$$\Omega = \{\Sigma, Pw - Q\}:P \subset \mathfrak{F}\{u_1, \cdots, u_q, w, y_1, \cdots, y_p\}.$$

Ω is prime, with arbitrary unknowns $u_1, \cdots, u_q$. Further,

$$\Sigma = \Omega \cap \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}.$$

Let

(6*) $$A, A_1, \cdots, A_p$$

be a basic set for Ω, introducing $w, y_1, \cdots, y_p$ in succession. We take, as we may, A algebraically irreducible.

We shall show that *each* A_i *is of order zero in* y_i *and is linear in* y_i. The equation $A = 0$ is called a *resolvent* of Σ.

Suppose that not every A_i is as described. Let A_j be the first A_i which either is of positive order or is not linear in y_i, and suppose that the order of A_j in y_j is r, with r positive.

Let U be the remainder of $ISI_1S_1 \cdots I_pS_pGP$ with respect to (6*), where the I's and J's are the initials and separants of the forms of the basic set (6*). Then U has the properties: 1. The order of U in w is not greater than that of A; 2. The order of U in y_i, $i = 1, \cdots, j$, is not greater than that of A_i; 3. U is free of $y_{j+1}, \cdots, y_p$; 4. $U \notin \Omega$.

Of all the forms in (U, Ω) which have the properties 1–4 listed for U, let V be one of least rank in y_j.

We shall prove that the order of V in y_j is less than r. Suppose it equals r. Then the degree of V in y_{jr} is no higher than that of A_j. Let J be the initial of V. Then $J \notin \Omega$.

For an appropriate integer m, we have

$$J^m A_j = EV + F,$$

where E is a nonzero form of lower degree than A_j in y_{jr}, and where F has the properties 1–3 and is of lower degree than V in y_{jr}. Hence $F \in \Omega$, so that $EV \in \Omega$, whence $E \in \Omega$.

Let $E = H_0 + H_1 y_{jr} + \cdots + H_t y_{jr}^t$, where $H_t \neq 0$, and each H_i is free of $y_{j+1}, \cdots, y_p$ and is of order less than r in y_j. The initial of $J^m A_j$ is identical with that of EV. Hence $H_t \notin \Omega$.

It is easy to see that there exist positive integers $a, a_1, \cdots, a_{j-1}$ such that

$$I^a I_1^{a_1} \cdots I_{j-1}^{a_{j-1}} H_i \equiv H_i' \qquad (\Omega), \qquad\qquad i = 0, 1, \cdots, t,$$

where each H_i' is reduced with respect to the ascending set $A, A_1, \cdots, A_{j-1}$. Since $H_t \notin \Omega$, we have $H_t' \neq 0$. Thus, $H_0' + H_1' y_{jr} + \cdots + H_t' y_{jr}^t$ is a nonzero form in Ω, reduced with respect to the basic set (6*). This contradiction shows that the order of V in y_j is less than r.

Now let

$$(7^*) \qquad\qquad \bar{u}_1, \cdots, \bar{u}_q, \bar{w}, \bar{y}_1, \cdots, \bar{y}_p$$

be a solution of Ω for which V (and consequently $ISI_1S_1 \cdots I_pS_pGP$) does not vanish. For $\bar{u}_1, \cdots, \bar{u}_q, \bar{w}, \bar{y}_1, \cdots, \bar{y}_{j-1}$, the forms $A_1, \cdots, A_{j-1}$ all vanish, A_j becomes a nonzero form $\bar{A}_j$ in y_j of order r, A_i, $i = j + 1, \cdots, p$, becomes a form $\bar{A}_i$ of order zero and linear in y_i, and V becomes a form $\bar{V}$ in y_j of order less than r. It follows, by Lemma 1*, that $(y_j - \bar{y}_j)\bar{V}$ does not hold $\bar{A}_j$.

Let $\tilde{y}_j$ be a solution of $\bar{A}_j$ for which $(y_j - \bar{y}_j)\bar{V}$ does not vanish. When $y_j = \tilde{y}_j$, $\bar{A}_i$, $i = j + 1, \cdots, p$, becomes a linear form of order zero in y_i alone. The new forms $\bar{A}_i$, $i = j + 1, \cdots, p$, may be solved to obtain a solution $\tilde{y}_{j+1}, \cdots, \tilde{y}_p$.

Then $\bar{u}_1, \cdots, \bar{u}_q, \bar{w}, \bar{y}_1, \cdots, \bar{y}_{j-1}, \tilde{y}_j, \cdots, \tilde{y}_p$ is a solution of Ω for which GP does not vanish, is distinct from (7*), but has the same $w = \bar{w}$ and the same $u_i = \bar{u}_i$, $i = 1, \cdots, q$. This contradicts the fundamental property of the triad G, P, Q, and shows that A_j is of order zero in y_j.

That A_j is of degree unity in y_j may be proved as in (R). This completes the proof.

The rest of Chapter 2 in (R), except for §33, may now be taken over, word for word.

Columbia University,
New York, N. Y.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

HANKEL AND OTHER EXTENSIONS OF DIRICHLET'S SERIES

BY R. E. GREENWOOD

(Received December 14, 1939)

1. Introduction

Series of the type

$$f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s), \tag{1.1}$$

where (a) s is the complex variable, $s = \sigma + it$, σ and t real;
(b) the numbers a_n are all complex constants;
(c) $\lambda_1 < \lambda_2 < \lambda_3 < \cdots < \lambda_n < \cdots$

$$\lim_{n \to \infty} \lambda_n = \infty$$

are known as Dirichlet series, and their theory has been considered by numerous investigators. Widder[1] has generalized these series by the consideration of a Stieltjes integral.

If we consider series of the type

$$g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s) \tag{1.2}$$

it may be possible to establish convergence properties, subject to certain restrictions on the function $G(z)$, for the series (1.2) similar to the well known convergence properties of the Dirichlet series (1.1).

In particular, this method of generalizing Dirichlet series was suggested from a consideration of series of the type

$$h(s) = \sum_{n=1}^{\infty} a_n (i\lambda_n s)^{\frac{1}{2}} H_\nu^{(1)}(i\lambda_n s), \tag{1.3}$$

where $H_\nu^{(1)}(z)$ is the Hankel function of the first kind of order ν and is related to the more familiar Bessel functions $J_\nu(z)$ and $J_{-\nu}(z)$ by the relation

$$H_\nu^{(1)}(z) = \frac{J_{-\nu}(z) - \exp(-\nu\pi i) J_\nu(z)}{i \sin \nu\pi}. \tag{1.4}$$

The function $H_\nu^{(1)}(z)$ has the integral representation[2] for $\Re(\nu) \geqq -\frac{1}{2}$

[1] D. V. Widder, "A Generalization of Dirichlet Series," Transactions of the American Mathematical Society, Vol. 31, p. 694f. (1929).

[2] G. N. Watson, "The Theory of Bessel Functions," page 168.

$$(1.5)\quad H_\nu^{(1)}(z) = \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} \frac{\exp\left[i\left(z - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\right]}{\Gamma(\nu + \frac{1}{2})} \int_0^\infty \exp(-u)u^{\nu - \frac{1}{2}}\left(1 + \frac{iu}{2z}\right)^{\nu - \frac{1}{2}} du.$$

If we replace z by iz and define

$$(1.6)\qquad M_\nu(z) = \int_0^\infty \exp(-u)u^{\nu-\frac{1}{2}}\left(1 + \frac{u}{2z}\right)^{\nu - \frac{1}{2}} du$$

$$(1.7)\qquad A_\nu = \frac{\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp\left[-i\left(\frac{\nu\pi}{2} + \frac{\pi}{4}\right)\right]}{\Gamma(\nu + \frac{1}{2})}$$

we see that

$$(1.8)\qquad (iz)^{\frac{1}{2}} H_\nu^{(1)}(iz) = A_\nu \exp(-z) M_\nu(z).$$

Therefore the series (1.3) can also be written as

$$(1.9)\qquad h(s) = \sum_{n=1}^{\infty} a_n A_\nu \exp(-\lambda_n s) M_\nu(\lambda_n s)$$

and thus is formally identifiable as a series of type (1.2).

For convenience we shall call series of type (1.2) modified exponential series, and the particular series of type (1.9) Hankel series. Moreover we shall suppose that we are given a fixed λ_n sequence satisfying

(a) $\qquad 0 < \lambda_1 < \lambda_2 < \lambda_3 < \cdots < \lambda_n < \cdots,$

(b) $\qquad \lim_{n \to \infty} \lambda_n = \infty.$

The symbol $K(a)$ will be used to indicate a positive constant, depending perhaps on a.

2. Lemmas Concerning the Function $G(z)$

In order to establish convergence properties for the modified exponential series (1.2) we shall assume that the function $G(z)$ satisfies the following properties (which properties are understood to be assumed whenever $G(z)$ is mentioned hereafter):

PROPERTY 1. *$G(z)$ is analytic for $\Re(z) > 0$.*

PROPERTY 2. *$G(z)$ is bounded for $\Re(z) \geqq z_0 > 0$, i.e. say*

$$|G(z)| < K(z_0) \quad \text{for} \quad \Re(z) \geqq z_0 > 0.$$

PROPERTY 3. *$G(z)$ has an asymptotic expansion for large z of the form*

$$(2.1)\qquad G(z) \cong b_0 + \frac{b_1}{z} + \frac{b_2}{z^2} + \cdots,$$

where $b_0 \neq 0$.

On the basis of these properties we can establish a number of lemmas, which we shall state for the function $G(xs)$ where x is a real variable.

LEMMA 1. *There exists a constant* τ *such that for*

(a) $$x \geqq \lambda_1 > 0,$$

(b) $$\Re(s) \geqq \tau > 0,$$

we have

$$| G(xs) | < K_1(\tau)^3.$$

PROOF: This is merely a restatement of property 2.

LEMMA 2. *Given* $\tau > 0$, *there exists an integer* $N_0(\tau)$ *such that for*

(a) $$x \geqq \lambda_n, \qquad n \geqq N_0(\tau),$$

(b) $$\Re(s) \geqq \tau > 0,$$

we have

$$0 < K_2(\tau) \leqq | G(xs) |.$$

PROOF: Using the asymptotic expansion

$$G(xs) \cong b_0 + \frac{b_1}{xs} + \frac{b_2}{x^2 s^2} + \cdots, \tag{2.1}$$

and noting that $b_0 \neq 0$, we can get

$$| G(xs) - b_0 | < \frac{B}{x\tau} \qquad \text{for large } x$$

where B is a constant, and hence we can clearly select x values and a constant $K_2(\tau)$ so that we obtain

$$0 < K_2(\tau) < | G(xs) |.$$

LEMMA 3. *Given* $\lambda_1 > 0$, *there exists a real number* $\xi(\lambda_1)$ *such that for*

(a) $$x \geqq \lambda_1 > 0,$$

(b) $$\Re(s) \geqq \xi > 0,$$

we have

$$0 < K_3(\tau) < | G(xs) |.$$

Proof is similar to that of lemma 2.

LEMMA 4. *For*

(a) $$\Re(s) \geqq \Re(s_0) = \tau > 0,$$

(b) $$x \geqq \lambda_n, \qquad n \geqq N_0(\tau),$$

[3] In reality K_1 depends also on λ_1, but we are considering the λ_n sequence as fixed and do not indicate this dependence.

we have

$$\left|\frac{G(xs)}{G(xs_0)}\right| \leqq K_4(\tau);$$

as follows at once from combining lemmas 1 and 2.

Now the asymptotic expansion for $dG(z)/dz$ can be obtained from the asymptotic expansion,

$$G(xs) \cong b_0 + \frac{b_1}{xs} + \frac{b_2}{x^2 s^2} + \cdots, \tag{2.1}$$

by termwise differentiation,[4] thereby leading to

$$\frac{\partial G(xs)}{\partial s} \cong -\frac{b_1}{xs^2} - \frac{2b_2}{x^2 s^3} - \cdots, \tag{2.2}$$

$$\frac{\partial G(xs)}{\partial x} \cong -\frac{b_1}{x^2 s} - \frac{2b_2}{x^3 s^2} - \cdots, \tag{2.3}$$

$$\frac{\partial^2 G(xs)}{\partial x \partial s} \cong \frac{b_1}{x^2 s^2} + \frac{4b_2}{x^3 s^3} + \cdots, \tag{2.4}$$

and from these expansions we obtain

LEMMA 5. *For $\Re(s) \geqq \tau > 0$ there exists an integer $N_1(\tau)$ such that for $x \geqq \lambda_n$, $n \geqq N_1(\tau)$ we have*

$$\left|\frac{\partial G(xs)}{\partial s}\right| < K_5(\tau),$$

$$\left|\frac{\partial G(xs)}{\partial x}\right| < K_6(\tau),$$

$$\left|\frac{\partial^2 G(xs)}{\partial x \partial s}\right| < K_7(\tau).$$

LEMMA 6. *For $\Re(s) \geqq \tau > 0$ there exists an integer $N_2(\tau)$ such that for $q \geqq p \geqq N_2(\tau)$ we have*

$$\int_{\lambda_p}^{\lambda_q} \left|\frac{\partial G(xs)}{\partial x}\right| dx \leqq K_8(\tau)\left[\frac{1}{\tau\lambda_p} - \frac{1}{\tau\lambda_q}\right] \leqq \frac{K_8(\tau)}{\tau\lambda_p}.$$

PROOF. This follows immediately from the asymptotic expansion (2.3). The inequality

$$\left|\frac{|f(z+h)| - |f(z)|}{h}\right| \leqq \frac{|f(z+h) - f(z)|}{|h|},$$

[4] Bochner, "Fourier Analysis," (planographed notes, Princeton University), pages 7 and 29; also Knopp, "Theorie und Anwendung der Unendlichen Reihen," third edition, p. 560.

shows that

$$\left|\frac{d|f(z)|}{dz}\right| \leqq \left|\frac{df(z)}{dz}\right|$$

and thus leads to

LEMMA 7. *For $\Re(s) \geqq \tau > 0$ and for $q \geqq p \geqq N_2(\tau)$ we have*

$$\left|\int_{\lambda_p}^{\lambda_q} \frac{\partial|G(xs)|}{\partial x}\,dx\right| \leqq \frac{K_8(\tau)}{\tau\lambda_p}.$$

LEMMA 8. *For $\Re(s) \geqq \Re(s_0) = \tau > 0$ and for $x \geqq \lambda_n$, $n \geqq \max\,[N_1(\tau), N_2(\tau)]$ we have*

$$\left|\frac{\dfrac{\partial G(xs_0)}{\partial x}}{G(xs_0)} - \frac{\dfrac{\partial G(xs)}{\partial x}}{G(xs)}\right| \leqq K_9(\tau)|s - s_0|.$$

PROOF. Writing the function as the integral of its own derivative, we have

$$\left|\frac{\dfrac{\partial G(xs_0)}{\partial x}}{G(xs_0)} - \frac{\dfrac{\partial G(xs)}{\partial x}}{G(xs)}\right| = \left|\int_{s_0}^{s} \frac{\dfrac{\partial^2 G(xz)}{\partial z\partial x}G(xz) - \dfrac{\partial G(xz)}{\partial x}\dfrac{\partial G(xz)}{\partial z}}{[G(xz)]^2}\,dz\right|$$

$$\leqq \left|\int_{s_0}^{s} \frac{K_7(\tau)K_1(\tau) + K_5(\tau)K_6(\tau)}{[K_2(\tau)]^2}\,dz\right| \leqq K_9(\tau)|s - s_0|$$

by applying lemmas 1, 2, and 5.

In particular, the function $M_\nu(z)$ defined by

$$(1.6) \qquad M_\nu(z) = \int_0^\infty \exp\,(-u)u^{\nu-\frac{1}{2}}\left(1 + \frac{u}{2z}\right)^{\nu-\frac{1}{2}} du,$$

satisfies Properties 1, 2, and 3. For $M_\nu(z)$ is analytic both in z and in ν, and has the asymptotic expansion[5]

$$M_\nu(z) \cong \sum_{m=0}^{\infty} \frac{\Gamma(\nu + \frac{1}{2})\Gamma(\nu + \frac{1}{2} + m)}{m!\,\Gamma(\nu + \frac{1}{2} - m)(2z)^m}$$

$$\cong \Gamma(\nu + \tfrac{1}{2})\left[1 + \sum_{m=1}^{\infty} \frac{\{4\nu^2 - 1^2\}\{4\nu^2 - 3^2\}\cdots\{4\nu^2 - (2m-1)^2\}}{2^{3m}m!\,z^m}\right]$$

and thus, since $b_0 \neq 0$, properties 1 and 3 are satisfied. Property 2 can be established by taking an upper bound to the complex valued integral representation (1.6) for $\Re(z) \geqq z_0 > 0$.

Thus any convergence proof for the modified exponential series

$$(1.2) \qquad g(s) = \sum_{n=1}^{\infty} a_n \exp\,(-\lambda_n s)G(\lambda_n s)$$

---

[5] Watson, l.c., p. 198, and Bochner, l.c., page 15. Care must be taken in interpreting the notations in these references for the case $m = 0$.

which uses only properties 1, 2, and 3 of the function $G(z)$ shows at the same time that the corresponding Hankel series

$$h(s) = \sum_{n=1}^{\infty} a_n A_\nu \exp(-\lambda_n s) M_\nu(\lambda_n s) \tag{1.9}$$

satisfies the same type of convergence.

For convenience we state a familiar lemma.

LEMMA 9. Abel's lemma on partial summation. *We have identically*

$$\sum_{n=p}^{n=q} U_n V_n \equiv \sum_{n=p}^{n=q-1} \left[(V_n - V_{n+1}) \sum_{m=p}^{m=n} U_m \right] + V_q \sum_{n=p}^{n=q} U_n .$$

3. Convergence of the Modified Exponential Series

THEOREM 1. *If the modified exponential series* (1.2) *is convergent for* $s_0 = \sigma_0 + it_0$ *where* $\sigma_0 > 0$, *then it is convergent for any* s *satisfying* $\Re(s) \geqq \Re(s_0) = \sigma_0$.

PROOF: The proof of this theorem is contained in the more general theorem which follows.

THEOREM 2. *If the modified exponential series* (1.2) *is convergent for* $s_0 = \sigma_0 + it_0$ *where* $\sigma_0 > 0$, *then it is uniformly convergent throughout the angular sector about the point* s_0 *in the* s *plane defined by the inequality*

$$|\operatorname{am}(s - s_0)| \leqq \psi < \frac{\pi}{2}.$$

PROOF: We need only show that for s in the angular sector, given $\delta > 0$, we can find an integer n_0 such that for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| < \delta.$$

In Abel's lemma on partial summation we identify

$$U_n = a_n \exp(-\lambda_n s_0) G(\lambda_n s_0),$$

$$V_n = \exp[-(s - s_0)\lambda_n] \frac{G(\lambda_n s)}{G(\lambda_n s_0)},$$

so that we have

$$\begin{aligned} \left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| &= \left| \sum_{n=p}^{n=q} U_n V_n \right| \\ &\leqq \sum_{n=p}^{n=q-1} \left| \exp[-(s - s_0)\lambda_n] \frac{G(\lambda_n s)}{G(\lambda_n s_0)} - \exp[-(s - s_0)\lambda_{n+1}] \frac{G(\lambda_{n+1} s)}{G(\lambda_{n+1} s_0)} \right| \\ &\qquad \cdot \left| \sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right| \\ &\quad + \left| \exp[-(s - s_0)\lambda_q] \frac{G(\lambda_q s)}{G(\lambda_q s_0)} \right| \cdot \left| \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right|. \end{aligned} \tag{3.1}$$

By hypothesis, given an $\epsilon > 0$, there exists an integer n_1 such that for $q \geqq p \geqq n_1$ we have

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0) \right| < \epsilon.$$

We take $n_0 = \max [N_0(\tau), N_1(\tau), n_1]$, where N_0 is given by lemma 4 and N_1 by lemma 5. Then (3.1) becomes

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right|$$

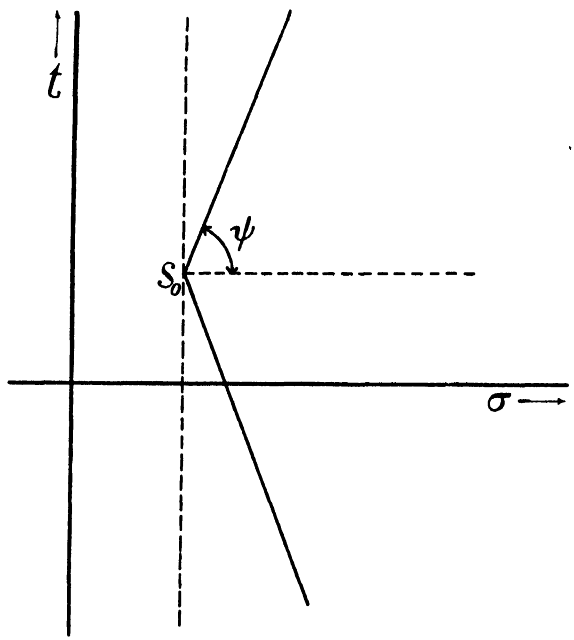


FIG. 1

$$(3.2) \qquad \leqq \epsilon \sum_{n=p}^{n=q-1} \left| \exp[-(s-s_0)\lambda_n] \frac{G(\lambda_n s)}{G(\lambda_n s_0)} - \exp[-(s-s_0)\lambda_{n+1}] \frac{G(\lambda_{n+1} s)}{G(\lambda_{n+1} s_0)} \right| + \epsilon \left| \exp[-(s-s_0)\lambda_q] \frac{G(\lambda_q s)}{G(\lambda_q s_0)} \right|.$$

Writing the function as the integral of its own derivative, we have for $n \geqq n_0$

$$\left| \exp[-(s-s_0)\lambda_n] \frac{G(\lambda_n s)}{G(\lambda_n s_0)} - \exp[-(s-s_0)\lambda_{n+1}] \frac{G(\lambda_{n+1} s)}{G(\lambda_{n+1} s_0)} \right|$$

$$\leqq \left| (s-s_0) \int_{\lambda_n}^{\lambda_{n+1}} \exp[-(s-s_0)x] \frac{G(xs)}{G(xs_0)} \, dx \right|$$

$$(3.3)\qquad + \left| \int_{\lambda_n}^{\lambda_{n+1}} \exp\left[-(s-s_0)x\right] \frac{G(xs)}{G(xs_0)} \left\{ \frac{\dfrac{\partial G(xs_0)}{\partial x}}{G(xs_0)} - \frac{\dfrac{\partial G(xs)}{\partial x}}{G(xs)} \right\} \right|$$

$$\leqq K_4(\sigma_0)\cdot|s-s_0|\cdot\left|\int_{\lambda_n}^{\lambda_{n+1}} \exp\left[-(s-s_0)x\right]dx\right|$$

$$+ K_4(\sigma_0)K_9(\sigma_0)\cdot|s-s_0|\cdot\left|\int_{\lambda_n}^{\lambda_{n+1}} \exp\left[-(s-s_0)x\right]dx\right|$$

by application of lemmas 4 and 8, and hence

$$\leqq K_4(\sigma_0)(1+K_9(\sigma_0))\frac{|s-s_0|}{\Re(s-s_0)}\left[\exp\left[-\Re(s-s_0)\lambda_n\right]-\exp\left[-\Re(s-s_0)\lambda_{n+1}\right]\right].$$

Substituting this inequality in (3.2) we get

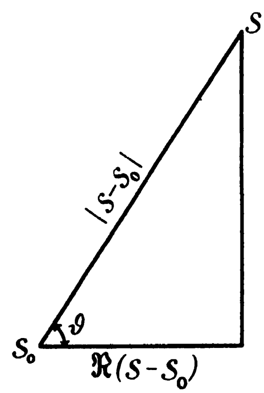


FIG. 2

$$\left|\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s)G(\lambda_n s)\right|$$

$$\leqq \frac{\epsilon|s-s_0|}{\Re(s-s_0)}K_4(\sigma_0)[1+K_9(\sigma_0)]\left[\exp\left[-\Re(s-s_0)\lambda_p\right]-\exp\left[-\Re(s-s_0)\lambda_q\right]\right.$$

$$(3.4)\qquad + \epsilon K_4(\sigma_0)\exp\left[-\Re(s-s_0)\lambda_q\right.$$

$$\leqq \frac{\epsilon|s-s_0|}{\Re(s-s_0)}K_4(\sigma_0)[1+K_9(\sigma_0)]\exp\left[-\Re(s-s_0)\lambda_p\right]$$

$$\leqq \epsilon K_{10}(\sigma_0)\frac{|s-s_0|}{\Re(s-s_0)}.$$

Now for am $(s-s_0)=\theta$, $|\theta|\leqq\psi<\frac{\pi}{2}$, we get

$$\frac{|s-s_0|}{\Re(s-s_0)}=\sec\theta\leqq\sec\psi,$$

and thus (3.4) becomes

$$\left|\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s)\right| \leqq \epsilon K_{10}(\sigma_0) \sec \psi,$$

which holds for all s in the angular sector and for $q \geqq p \geqq n_0$, and thus the required condition is satisfied if we specify that we choose $\epsilon = \delta/(K_{10}(\sigma_0) \sec \psi)$ and the proof is complete.

Definition: We suppose that the modified exponential series is not identically zero. Then, as is the case with Dirichlet series, the series (1.2) may:

(a) Converge for all s, $\Re(s) > 0$

(b) Converge for no s, $\Re(s) > 0$

(c) Converge for some values of s, $\Re(s) > 0$.

In the latter case there is a real number α such that the series is convergent for all s satisfying $\Re(s) > \alpha$, and divergent or oscillatory for all s satisfying $0 < \Re(s) < \alpha$. The number α is called the *abscissa of convergence* of the modified exponential series, and the line defined by $\Re(s) = \alpha$ is called the *line of convergence*. The symbol α will be reserved for the abscissa of convergence of the modified exponential series (1.2). The remarks of Hardy and Riesz[6] relative to the convergence of a Dirichlet series on the line of convergence are applicable also to modified exponential series. In section 7 of this paper we consider a special case where the line of convergence for the Hankel series (1.9) contains at least one singular point.

We shall now prove two theorems which will show equi-convergence of the Dirichlet series and the modified exponential series.

Theorem 3. *If the Dirichlet series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ *converges for the point* $s_0 = \sigma_0 + it_0$, $\sigma_0 > 0$, *then the modified exponential series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0)$ *is also convergent.*

Proof: In Abel's lemma we identify

$$U_n = a_n \exp(-\lambda_n s_0),$$
$$V_n = G(\lambda_n s_0),$$

and thus we get

$$\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0)$$

$$(3.5)\qquad \equiv \sum_{n=p}^{n=q-1} [G(\lambda_n s_0) - G(\lambda_{n+1} s_0)] \left[\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0)\right] + G(\lambda_q s_0) \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s_0).$$

[6] Hardy and Riesz, "The General Theory of Dirichlet's Series," page 9.

Now by hypothesis, given $\epsilon > 0$, there exists an integer n_0 such that for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) \right| < \epsilon.$$

Now choose $n_1 = \max[n_0, N_2(\sigma_0)]$ where N_2 is given by lemma 6. We get for $q \geqq p \geqq n_1$

$$\left| \sum_{n=p}^{n=q-1} [G(\lambda_n s_0) - G(\lambda_{n+1} s_0)] \left[\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) \right] \right|$$

$$\leqq \sum_{n=p}^{n=q-1} \left| \sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) \right| \int_{\lambda_n}^{\lambda_{n+1}} \left| \frac{\partial G(xs_0)}{\partial x} \right| dx \tag{3.6}$$

$$\leqq \epsilon \int_{\lambda_p}^{\lambda_q} \left| \frac{\partial G(xs_0)}{\partial x} \right| dx \leqq \frac{\epsilon K_8(\sigma_0)}{\sigma_0 \lambda_p} \leqq \frac{\epsilon K_8(\sigma_0)}{\sigma_0 \lambda_1}$$

on application of lemma 6.

On reducing (3.5) by use of (3.6) and lemma 1, we get, for $q \geqq p \geqq n_1$

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0) \right| \leqq \frac{\epsilon K_8(\sigma_0)}{\sigma_0 \lambda_1} + K_1(\sigma_0)\epsilon \leqq \epsilon K(\sigma_0),$$

and thus the convergence of the modified exponential series is established.

THEOREM 4. *If the modified exponential series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *converges for the point* $s_0 = \sigma_0 + it_0$, $\sigma_0 > 0$, *then the Dirichlet series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_0)$ *is also convergent.*

PROOF. In Abel's lemma we identify

$$U_n = a_n \exp(-\lambda_n s_0) G(\lambda_n s_0),$$

$$V_n = \frac{1}{G(\lambda_n s_0)}.$$

Now, by hypothesis, given $\epsilon > 0$, there exists an integer n_0 such that for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0) \right| < \epsilon.$$

Now choose $n_1 = \max[N_0(\sigma_0), N_2(\sigma_0), n_0]$ where N_0 and N_2 are given by lemmas 2 and 6 respectively. Similar to the proof of the previous theorem, we get, for $q \geqq p \geqq n_1$

$$\left| \sum_{n=p}^{n=q-1} \left[\frac{1}{G(\lambda_n s_0)} - \frac{1}{G(\lambda_{n+1} s_0)} \right] \left[\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right] \right|$$

$$(3.7) \qquad \leqq \sum_{n=p}^{n=q-1} \left| \sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right| \int_{\lambda_n}^{\lambda_{n+1}} \left| \frac{\partial \left(-\frac{1}{G(x s_0)} \right)}{\partial x} \right| dx$$

$$\leqq \epsilon \int_{\lambda_p}^{\lambda_q} \left| \frac{\frac{\partial G(x s_0)}{\partial x}}{[G(x s_0)]^2} \right| dx \leqq \frac{\epsilon K_8(\sigma_0)}{\sigma_0 \lambda_p [K_2(\sigma_0)]^2}$$

by application of lemma 2 and lemma 6.

Abel's lemma then takes the form

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) \right|$$

$$\leqq \left| \sum_{n=p}^{n=q-1} \left[\frac{1}{G(\lambda_n s_0)} - \frac{1}{G(\lambda_{n+1} s_0)} \right] \left[\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right] \right|$$

$$+ \left| \frac{1}{G(\lambda_q s_0)} \right| \cdot \left| \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right|$$

$$\leqq \frac{\epsilon K_8(\sigma_0)}{\sigma_0 \lambda_p [K_2(\sigma_0)]^2} + \frac{\epsilon}{K_2(\sigma_0)} \leqq \epsilon K(\sigma_0)$$

for $q \geqq p \geqq n_1$, and thus the convergence of the Dirichlet series is established.

4. Absolute Convergence of the Modified Exponential Series

THEOREM 5. *If the modified exponential series* (1.2) *is absolutely convergent for* $s_1 = \sigma_1 + it_1$ *where* $\sigma_1 > 0$, *then it is absolutely convergent in the half plane* $\Re(s) \geqq \sigma_1$.

PROOF. We need only show that for s in the half plane $\Re(s) \geqq \sigma_1$, given $\delta > 0$, we can find an integer n_0 such that for $q \geqq p \geqq n_0$ we have

$$\sum_{n=p}^{n=q} |a_n| \exp(-\lambda_n \sigma) |G(\lambda_n s)| < \delta.$$

But by hypothesis, given $\epsilon > 0$, there exists an integer n_1 such that for $q \geqq p \geqq n_1$ we have

$$\sum_{n=p}^{n=q} |a_n| \exp(-\lambda_n \sigma_1) |G(\lambda_n s_1)| < \epsilon.$$

We take $n_0 = \max [n_1, N_0(\sigma_1)]$ where N_0 is given by lemma 2. Then for $q \geqq p \geqq n_0$ we get

$$\sum_{n=p}^{n=q} |a_n| \exp(-\lambda_n \sigma) |G(\lambda_n s)|$$

$$\leqq \sum_{n=p}^{n=q} |a_n| \exp(-\lambda_n \sigma_1) |G(\lambda_n s_1)| \left| \frac{G(\lambda_n s)}{G(\lambda_n s_1)} \right| \exp [(\sigma_1 - \sigma)\lambda_n]$$

$$\leqq \epsilon K_4(\sigma_1)$$

by use of the above and of lemma 4, and thus the proof is completed.

Similar to the case of convergence, there may exist a half plane of absolute convergence for the modified exponential series, and β will be used to denote the *abscissa of absolute convergence* and the equation $\Re(s) = \beta$ will be used to denote the *line of absolute convergence.*

The following two theorems will show equi-absolute convergence of the Dirichlet series and the modified exponential series.

THEOREM 6. *If the Dirichlet series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_1)$ *is absolutely convergent and* $\Re(s_1) = \sigma_1 > 0$, *then the modified exponential series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_1) G(\lambda_n s_1)$ *is also absolutely convergent.*

PROOF. The proof is essentially a repetition of the argument of theorem 3 except that lemma 7 must be used instead of lemma 6.

THEOREM 7. *If the modified exponential series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *is absolutely convergent for* $s = s_1 = \sigma_1 + it_1$, *where* $\sigma_1 > 0$, *then the Dirichlet series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_1)$ *is also absolutely convergent.*

PROOF. The proof follows that of theorem 4 except that lemma 7 must be used instead of lemma 6.

5. Uniform Convergence of the Modified Exponential Series

THEOREM 8. *If the modified exponential series* $g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *is absolutely convergent for* $s = s_1$, $\Re(s_1) = \beta > 0$, *then the series is uniformly convergent for* $\Re(s) \geqq \beta$.

PROOF. The method of proof for theorem 5 establishes also the above theorem.

However, it may happen that the series is uniformly convergent for a certain range of values of s for which it is not absolutely convergent. Indeed, the series will in general have a half plane of uniform convergence, and we use γ to denote the *abscissa of uniform convergence.* It follows at once that the three abscissae of convergence satisfy the inequalities $\alpha \leqq \gamma \leqq \beta$.

Furthermore, the Dirichlet series and the modified exponential series are equi-uniformly convergent as is shown by the following two theorems.

THEOREM 9. *If the Dirichlet series* $f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ *is uniformly convergent for* $\Re(s) \geqq \gamma > 0$, *then the modified exponential series* $g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *is also uniformly convergent for* $\Re(s) \geqq \gamma > 0$.

PROOF. In Abel's lemma we identify

$$U_n = a_n \exp(-\lambda_n s)$$

$$V_n = G(\lambda_n s).$$

Now, by hypothesis, given $\epsilon > 0$, there exists an integer n_0 such that for $\Re(s) \geqq \gamma$ and for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s) \right| < \epsilon.$$

Thus for $q \geqq p \geqq n_0$ and for $\Re(s) \geqq \gamma$ we get

$$\begin{aligned} \left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| &\leqq \left| \sum_{n=p}^{n=q-1} (G(\lambda_n s) - G(\lambda_{n+1} s)) \sum_{m=p}^{m=n} a_m \exp(-\lambda_m s) \right| \\ &\qquad + |G(\lambda_q s)| \left| \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s) \right| \\ &\leqq \epsilon \left[\sum_{n=p}^{n=q-1} |G(\lambda_n s) - G(\lambda_{n+1} s)| + |G(\lambda_q s)| \right]. \end{aligned}$$

Now choose $n_1 = \max [n_0, N_2(\gamma)]$ where N_2 is given by lemma 6. Then by use of lemmas 1 and 6, for $q \geqq p \geqq n_1$ and for all s satisfying $\Re(s) \geqq \gamma$ we get

$$\begin{aligned} \left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| &\leqq \epsilon \left[\sum_{n=p}^{n=q-1} \left| \int_{\lambda_n}^{\lambda_{n+1}} -\frac{\partial G(xs)}{\partial x} dx \right| + |G(\lambda_q s)| \right] . \\ &\leqq \epsilon \left[\int_{\lambda_p}^{\lambda_q} \left| \frac{\partial G(xs)}{\partial x} \right| dx + |G(\lambda_q s)| \right] \\ &\leqq \epsilon \left[\frac{K_8(\gamma)}{|s| \lambda_p} + K_1(\gamma) \right] \leqq \epsilon \left[\frac{K_8(\gamma)}{\gamma \lambda_1} + K_1(\gamma) \right] \\ &\leqq \epsilon K(\gamma) \end{aligned}$$

thereby completing the proof.

THEOREM 10. *If the modified exponential series* $g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *is uniformly convergent for* $\Re(s) \geqq \gamma > 0$, *then the Dirichlet series* $f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ *is uniformly convergent for* $\Re(s) \geqq \gamma > 0$.

PROOF. In Abel's lemma we identify

$$U_n = a_n \exp(-\lambda_n s) G(\lambda_n s),$$

$$V_n = \frac{1}{G(\lambda_n s)}.$$

Now by hypothesis, given $\epsilon > 0$, there exists an integer n_0 such that for all s satisfying $\Re(s) \geqq \gamma > 0$ and for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| < \epsilon.$$

Therefore for $q \geqq p \geqq n_0$ and for all s satisfying $\Re(s) \geqq \gamma > 0$ we get

$$\left|\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s)\right| \leqq \left|\sum_{n=p}^{n=q-1}\left(\frac{1}{G(\lambda_n s)} - \frac{1}{G(\lambda_{n+1} s)}\right)\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s)G(\lambda_m s)\right|$$
$$+ \left|\frac{1}{G(\lambda_q s)}\right|\left|\sum_{m=p}^{m=q} a_m \exp(-\lambda_m s)G(\lambda_m s)\right|$$
$$\leqq \epsilon\left[\sum_{n=p}^{n=q-1}\left|\frac{1}{G(\lambda_n s)} - \frac{1}{G(\lambda_{n+1} s)}\right| + \left|\frac{1}{G(\lambda_q s)}\right|\right].$$

Now choose $n_1 = \max[n_0, N_0(\gamma), N_2(\gamma)]$ where N_0 and N_2 are given by lemmas 2 and 6. Then, by the use of these lemmas, for $q \geqq p \geqq n_1$, and for all s satisfying $\Re(s) \geqq \gamma$ we get

$$\left|\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s)\right| \leqq \left[\sum_{n=p}^{n=q-1}\left|\int_{\lambda_n}^{\lambda_{n+1}} \frac{\dfrac{\partial G(xs)}{\partial x}}{[G(xs)]^2}\,dx\right| + \left|\frac{1}{G(\lambda_q s)}\right|\right]$$
$$\leqq \epsilon\left[\frac{K_8(\gamma)}{|s|\,[K_2(\gamma)]^2\lambda_p} + \frac{1}{K_2(\gamma)}\right] \leqq \epsilon\left[\frac{K_8(\gamma)}{\gamma\lambda_1[K_2(\gamma)]^2} + \frac{1}{K_2(\gamma)}\right]$$
$$\leqq \epsilon K(\gamma),$$

thereby completing the proof.

6. Consequences of Convergence

As a result of theorem 2 a number of properties of the modified exponential series are readily deduced. We state a few of these.

THEOREM 11. *If D is any finite region in the s plane such that for all points of D we have $\sigma \geqq \alpha + \delta > \alpha > 0$, then the modified exponential series (1.2) is uniformly convergent throughout $\bar{D}$, and its sum, the function $g(s)$, represents an analytic function in D.*

PROOF. This is an immediate consequence of theorem 2 and Weierstrass's classical theorem about uniformly convergent series of analytic functions.[7]

COROLLARY. *The derivative of $g(s)$ may be computed by termwise differentiation.*

THEOREM 12. *If the modified exponential series is convergent for the point $s_0 = \sigma_0 + it_0$ (where $\sigma_0 > 0$) to the value $g(s_0)$, then $g(s) \to g(s_0)$ whenever $s \to s_0$ along any path which lies in the angular sector* $|\text{am}\,(s - s_0)| \leqq \psi < \pi/2$.

PROOF. This is an immediate consequence of theorem 2.

THEOREM 13. *If the modified exponential series $g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)G(\lambda_n s)$ is convergent for $s = s_0 = \sigma_0 + it_0$ where $\sigma_0 > 0$, and if E denotes the angular sector*

$$|\text{am}\,(s - s_0)| \leqq \psi < \frac{\pi}{2}$$

[7] Titchmarsh, "The Theory of Functions," page 95.

and if $g(s) = 0$ *for an infinity of values of* s *lying in* E, *then* $a_n = 0$ *for all values of* n.

PROOF. We assume $g(s) \not\equiv 0$, else the proof is trivial. Since $g(s)$ is analytic in E, in any finite neighborhood of a point of E there must be only a finite

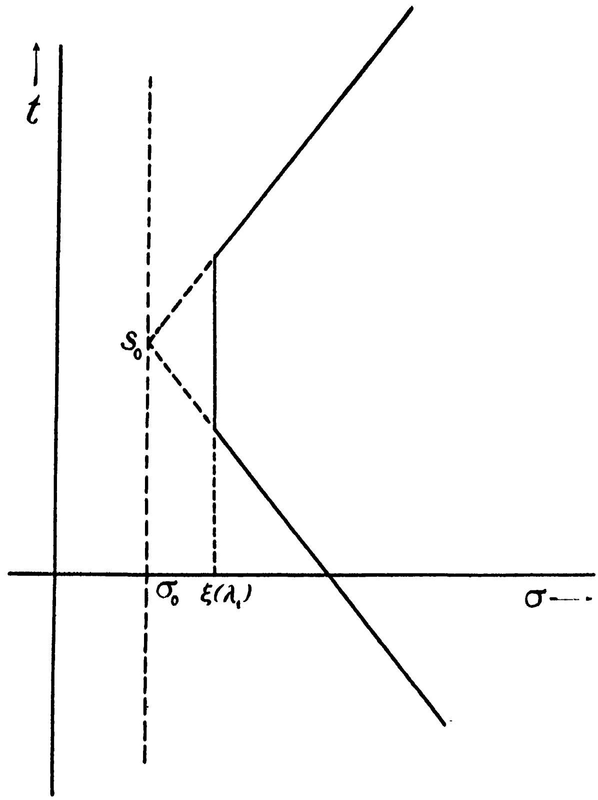


FIG. 3

number of zeros of $g(s)$. Because of the angular character of the set E it must be possible to select a sequence of values $s_n = \sigma_n + it_n$ where

(a) $\sigma_n > \xi(\lambda_1)$, ξ given by lemma 3

(b) $\lim_{n \to \infty} \sigma_n = \infty$

and such that $g(s_n) = 0$ for each n.

We consider the function

$$r(s) = \frac{\exp(\lambda_1 s)}{G(\lambda_1 s)} g(s)$$

$$= a_1 + \sum_{n=2}^{\infty} a_n \exp[-(\lambda_n - \lambda_1)s] \frac{G(\lambda_n s)}{G(\lambda_1 s)}.$$

This new function $r(s)$ is convergent for $s = s_0$ and is uniformly convergent in E' where E' denotes the deleted (if $\xi(\lambda_1) > \sigma_0$) angular sector

$$\sigma \geqq \max [\xi(\lambda_1), \sigma_0] > 0$$

$$|\operatorname{am} (s - s_0)| \leqq \psi < \frac{\pi}{2}.$$

We let $s \to \infty$ along any set of values whatsoever in E'. Since for a uniformly convergent series we can take the termwise limit,[8] we get $r(s) \to a_1$. But if we let $s \to \infty$ over the set $\{s_n\}$ we get $r(s) \to 0$. Thus $a_1 = 0$. Since this argument may be repeated for each n, the theorem follows at once.[9]

THEOREM 14. *If the two series*

$$\psi(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$$

$$\varphi(s) = \sum_{n=1}^{\infty} b_n \exp(-\lambda_n s) G(\lambda_n s)$$

are convergent at $s = s_0 = \sigma_0 + it_0$ where $\sigma_0 > 0$ and if in an angular sector E there are an infinity of values s_n such that $\psi(s_n) = \varphi(s_n)$ then $a_n = b_n$ for all n.

PROOF. On writing $g(s) = \psi(s) - \varphi(s)$, we see that the previous theorem applies, and thus we get our conclusion.

As a consequence of the sets of equi-convergence theorems, we have the following results:[10]

THEOREM 15. *If the abscissa of convergence α of the modified exponential series*

$$g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$$

is positive, then it is given by the formula

$$\alpha = \overline{\lim_{n \to \infty}} \frac{\log \left| \sum_{m=1}^{n} a_m \right|}{\lambda_n}.$$

THEOREM 16. *If the abscissa of absolute convergence β of the modified exponential series is positive, then it is given by*

$$\beta = \overline{\lim_{n \to \infty}} \frac{\log \sum_{m=1}^{n} |a_m|}{\lambda_n}.$$

[8] Titchmarsh, l.c., page 8.

[9] This proof is essentially borrowed from Hardy and Riesz, l.c., page 6.

[10] Hardy and Riesz, l.c., pages 7–9.

THEOREM 17. *We have*

$$\beta - \alpha \leqq \overline{\lim_{n\to\infty}} \frac{\log n}{\lambda_n}.$$

7. Convergence on the Line of Convergence for the Hankel Series

For Dirichlet series it is well known that under certain special conditions the real point of the line of convergence is a singularity of the function.[11] The following analogous theorem holds for the Hankel series (1.9).

THEOREM 18. *If ν is real, $\nu \geqq \frac{1}{2}$, and if all the coefficients of the Hankel series are positive or zero, and if $\Re(s) = \alpha > 0$ defines the line of convergence, then the point $s = \alpha$ is a singular point of the function represented by the series.*

PROOF. Let the series be represented by

$$\begin{aligned} h(s) &= \sum_{n=1}^{\infty} a_n (i\lambda_n s)^{\frac{1}{2}} H_\nu^{(1)}(i\lambda_n s) \\ &= \sum_{n=1}^{\infty} a_n A_\nu \exp(-\lambda_n s) M_\nu(\lambda_n s). \end{aligned}$$

It will be convenient to use the function $K_\nu(z)$ defined by[12]

$$K_\nu(z) = \frac{i\pi}{2} \exp\left(\frac{\nu\pi i}{2}\right) H_\nu^{(1)}(iz). \tag{7.1}$$

This function has the integral representation[13]

$$K_\nu(z) = \int_0^\infty \exp(-z \cosh t) \cosh \nu t \, dt. \tag{7.2}$$

Thus we consider the function

$$h(s) = (i)^{\frac{1}{2}} \frac{2}{\pi i} \exp\left(-\frac{\nu\pi i}{2}\right) \sum_{n=1}^{\infty} a_n (\lambda_n s)^{\frac{1}{2}} K_\nu(\lambda_n s) \tag{7.3}$$

which is analytic at the point $s = \alpha + 1 = \alpha_1$, and hence can be expanded in a Taylor's series about this point. We shall assume that the point $s = \alpha$ is not a singular point of the function $h(s)$, then the above Taylor's series expansion must be valid for a value of s smaller than α, say $\alpha_2 < \alpha$. By justifying certain summation interchanges we shall show that this requires $h(s)$ to be convergent at the point $s = \alpha_2 < \alpha$, thereby establishing a contradiction.

The domain in which we are interested does not contain the origin, so that the function

$$r(s) = \frac{\pi i}{2(i)^{\frac{1}{2}}} \exp\left(\frac{\nu\pi i}{2}\right) \frac{h(s)}{(s)^{\frac{1}{2}}} \tag{7.4}$$

[11] Hardy and Riesz, l.c., page 10.
[12] Watson, l.c., page 78.
[13] Watson, l.c., page 181.

will have the same singularities as $h(s)$. The Taylor's series expansion of the function $r(s)$ about the point $s = \alpha_1$ will, by hypothesis, converge at the point $s = \alpha_2$. Indeed

$$r(s) = \frac{\pi i}{2(i)^{\frac{1}{2}}} \exp\left(\frac{\nu\pi i}{2}\right) \frac{h(s)}{(s)^{\frac{1}{2}}}$$
$$= \sum_{n=1}^{\infty} a_n(\lambda_n)^{\frac{1}{2}} \int_0^{\infty} \exp(-\lambda_n s \cosh t) \cosh \nu t \, dt$$

and

$$r(s) = \sum_{m=0}^{\infty} \frac{(s-\alpha_1)^m}{m!} r^{(m)}(\alpha_1)$$
$$= \sum_{m=0}^{\infty} \frac{(s-\alpha_1)^m}{m!} \sum_{n=1}^{\infty} a_n(\lambda_n)^{\frac{1}{2}}(-1)^m(\lambda_n)^m$$
$$\cdot \int_0^{\infty} \exp(-\alpha_1\lambda_n \cosh t)(\cosh t)^m \cosh \nu t \, dt.$$

We note that the value of the integral is positive since the integrand is always positive, that $(-1)^m(\alpha_2 - \alpha_1)^m$ is positive, and we recall that we assumed the coefficients a_n were all positive. Hence we can interchange the order of summation[14] and we have

$$r(\alpha_2) = \sum_{n=1}^{\infty} a_n(\lambda_n)^{\frac{1}{2}} \sum_{m=0}^{\infty} \frac{(\alpha_1 - \alpha_2)^m}{m!} \int_0^{\infty} \exp(-\alpha_1\lambda_n \cosh t)(\lambda_n \cosh t)^m \cosh \nu t \, dt.$$

We recognize that

$$\sum_{m=0}^{\infty} \frac{(\alpha_1 - \alpha_2)^m}{m!} \int_0^{\infty} \exp(-\alpha_1\lambda_n \cosh t)(\lambda_n \cosh t)^m \cosh \nu t \, dt$$

is the Taylor's series expansion for $\int_0^{\infty} \exp(-\alpha_2\lambda_n \cosh t) \cosh \nu t \, dt$ about the point α_1.

Thus we have that $r(\alpha_2)$ converges,

$$r(\alpha_2) = \sum_{n=1}^{\infty} a_n\sqrt{\lambda_n} \int_0^{\infty} \exp(-\alpha_2\lambda_n \cosh t) \cosh \nu t \, dt$$
$$= \sum_{n=1}^{\infty} a_n\sqrt{\lambda_n}\, K_\nu(\lambda_n \alpha_2).$$

Therefore the series

$$h(\alpha_2) = \frac{2\sqrt{i}}{\pi i} \exp\left(\frac{-\nu\pi i}{2}\right) \sqrt{\alpha_2}\, r(\alpha_2) = \sum_{n=1}^{\infty} a_n\sqrt{i\lambda_n s}\, H_\nu^{(1)}(i\lambda_n s)$$
$$= \sum_{n=1}^{\infty} a_n A_\nu \exp(-\lambda_n s) M_\nu(\lambda_n s)$$

[14] Bromwich, "Infinite Series," page 78.

converges, and since $\alpha_2 < \alpha$ we have a contradiction, α being defined as the abscissa of convergence. Hence the value $s = \alpha$ must be a singular point for the function $h(s)$.

8. Inversion Formulae

For the Dirichlet series $f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ Perron has given the following formula for the sum of the coefficients

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{f(s)}{s} \exp(\omega s)\, ds = \sum_{m=1}^{n} a_m, \tag{8.1}$$

where

$$\lambda_n < \omega < \lambda_{n+1},$$
$$c > \alpha.$$

In order to establish related formulae for the modified exponential series, we shall first prove a theorem concerning the order of the function $g(s)$ as $t \to \infty$.

THEOREM 19. *If the series*

$$g(s) = \sum_{m=1}^{\infty} a_m \exp[-(\lambda_m - h)s]G(\lambda_m s),$$

where s is the complex variable $s = \sigma + it$ and where $0 \leqq h \leqq \lambda_1$, is convergent or finitely oscillating at the point $s = \sigma_0 + 0$, then given $\epsilon > 0$, $\delta > 0$ we can find a value t_0 such that for $|t| \geqq t_0$, $\sigma \geqq \sigma_0 + \epsilon$, and for all n we have

$$\left| \frac{1}{t} \sum_{1}^{n} a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) \right| < \delta, \tag{8.2}$$

or

$$\sum_{1}^{n} a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) = o(|t|).^{15} \tag{8.3}$$

In particular, for $n = \infty$, we have

$$g(s) = o(|t|) \tag{8.4}$$

which holds uniformly for $\sigma \geqq \sigma_0 + \epsilon$.

PROOF. Because of the convergence or finite oscillation of the series at $s = \sigma_0$, and because of the fact that there are only a finite number of terms of type (a) below, there exists a constant L such that for all integers m, p, and $q \geqq p$, and for $r \leqq \max[N_0(\sigma_0), N_1(\sigma_0)]$ where N_0 and N_1 are given by lemmas 2 and 5 we obtain

$$|a_r \exp[-(\lambda_r - h)\sigma_0]| < L \tag{a}$$

[15] The notation $F(x) = o(\varphi(x))$ means that as x approaches a definite limit, $\lim \dfrac{F(x)}{\varphi(x)} = 0$.

(b) $$| a_m \exp [-(\lambda_m - h)\sigma_0] G(\lambda_m \sigma_0) | < L$$

(c) $$\left| \sum_{m=p}^{q} a_m \exp [-(\lambda_m - h)\sigma_0] G(\lambda_m \sigma_0) \right| < L.$$

Let N be a function of $| t |$, say $N(| t |)$. We consider sums of the form

$$\sum_1^n a_m \exp [-(\lambda_m - h)s] G(\lambda_m s)$$

and treat two cases, depending on whether $N < n$ or $N \geqq n$.

Case 1. Let

$$1 \leqq \max [N_0, N_1] < N < n. \tag{8.5}$$

We write

$$\sum_1^n a_m \exp [-(\lambda_m - h)s] G(\lambda_m s) = \sum_1^{N-1} + \sum_N^n.$$

Apply Abel's lemma to the summation $\sum_N^n$ to get

$$\begin{aligned} &\sum_1^n a_m \exp [-(\lambda_m - h)s] G(\lambda_m s) \\ &\quad = \sum_1^{N-1} a_m \exp [-(\lambda_m - h)s] G(\lambda_m s) + \sum_{m=N}^{n-1} \Bigg[\exp [-(s - \sigma_0)(\lambda_m - h)] \frac{G(\lambda_m s)}{G(\lambda_m \sigma_0)} \\ &\qquad\qquad - \exp [-(s - \sigma_0)(\lambda_{m+1} - h)] \frac{G(\lambda_{m+1} s)}{G(\lambda_{m+1} \sigma_0)} \Bigg] \\ &\qquad\qquad \cdot \left[\sum_{p=N}^{p=m} a_p \exp [-(\lambda_p - h)\sigma_0] G(\lambda_p \sigma_0) \right] \\ &\quad + \left[\sum_{m=N}^{m=n} a_m \exp [-(\lambda_m - h)\sigma_0] G(\lambda_m \sigma_0) \right] \exp [-(s - \sigma_0)(\lambda_n - h)] \frac{G(\lambda_n s)}{G(\lambda_n \sigma_0)} \\ &\quad = S_1 + S_2 + S_3. \end{aligned} \tag{8.6}$$

We have

$$\begin{aligned} | S_1 | &= \left| \sum_1^{N-1} a_m \exp [-(\lambda_m - h)s] G(\lambda_m s) \right| \\ &\leqq \sum_1^{N-1} | a_m \exp [-(\lambda_m - h)\sigma_0] | \cdot K_1(\sigma_0) \\ &\leqq (N - 1) \cdot L \cdot K_1 \end{aligned} \tag{8.7}$$

by use of lemma 1 and inequality (a). Also

$$
\begin{aligned}
|S_3| &= \left| \sum_{m=N}^{n} a_m \exp[-(\lambda_m - h)\sigma_0] G(\lambda_m \sigma_0) \right| \\
&\qquad \cdot \left| \exp[-(s - \sigma_0)(\lambda_n - h)] \frac{G(\lambda_n s)}{G(\lambda_n \sigma_0)} \right| \qquad (8.8) \\
&\leqq L \cdot K_4(\sigma_0) \\
&\leqq L \cdot K_4
\end{aligned}
$$

by use of lemma 4 and inequality (c).

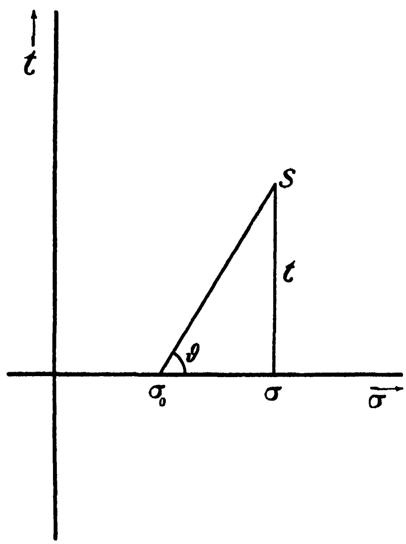


FIG. 4

Now the sum

$$
S_2 = \sum_{m=N}^{n-1} \left[\exp[-(\lambda_m - h)(s - \sigma_0)] \frac{G(\lambda_m s)}{G(\lambda_m \sigma_0)} - \exp[-(\lambda_{m+1} - h)(s - \sigma_0)] \frac{G(\lambda_{m+1} s)}{G(\lambda_{m+1} \sigma_0)} \right] \left[\sum_{p=N}^{m} a_p \exp[-(\lambda_p - h)\sigma_0] G(\lambda_p \sigma_0) \right]
$$

save for the factor $\exp[h(s - \sigma_0)]$ can be dealt with in the same way as was done in the proof of theorem 2, and similar to relation (3.4) we get

$$
|S_2| \leqq L \cdot K_{10}(\sigma_0) \frac{|s - \sigma_0|}{\Re(s - \sigma_0)} \exp[-(\lambda_N - h)\Re(s - \sigma_0)].
$$

Now suppose am $(s - \sigma_0) = \theta$. Then

$$
\frac{|s - \sigma_0|}{\Re(s - \sigma_0)} = \sec\theta = \sqrt{1 + \tan^2\theta} \leqq \sqrt{1 + \frac{t^2}{\epsilon^2}}
$$

for $\sigma \geqq \sigma_0 + \epsilon$. Thus we have

$$|S_2| \leqq L \cdot K_{10} \cdot \sqrt{1 + \frac{t^2}{\epsilon^2}} \exp[-(\lambda_N - h)\epsilon]$$

or

$$S_2 = O(|t \exp[-(\lambda_N - h)\epsilon]|).[16] \tag{8.9}$$

Therefore, substituting (8.7), (8.8), and (8.9) in relation (8.6) we get for $1 \leqq \max[N_0, N_1] < N < n$

$$\left|\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s)\right| = |S_1 + S_2 + S_3|$$

$$\leqq (N - 1) \cdot L \cdot K_1 + L \cdot K_{10} \sqrt{1 + \frac{t^2}{\epsilon^2}} \exp[-(\lambda_N - h)\epsilon] + L \cdot K_4$$

or

$$\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) = O(N) + O(|t \exp[-(\lambda_N - h)\epsilon]|). \tag{8.10}$$

Case 2. For $n \leqq N$ we get at once

$$\left|\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s)\right| \leqq nL \cdot K_1 \leqq N \cdot L \cdot K_1$$

or

$$\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) = O(N). \tag{8.11}$$

Now let $N(|t|)$ be so chosen that $N(|t|)$ tends to infinity more slowly than $|t|$, and also so that relation (8.5) is satisfied. Then for either of the two cases for $\Re(s) = \sigma \geqq \sigma_0 + \epsilon$ and for all n we have

$$\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) = o(|t|), \tag{8.12}$$

thereby completing the proof.

Lemma 10. *If* $b > 0$, $k \geqq 0$, then

$$\frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \exp(\omega s) \frac{ds}{s^{k+1}} = \frac{\omega^k}{\Gamma(k+1)}, \qquad \omega \geqq 0,$$

$$= 0, \qquad \omega < 0.$$

See Hardy and Riesz for proof.[17]

---

[16] The notation $F(x) = O(\varphi(x))$ means that $|F(x)| < K\varphi(x)$ for values of x sufficiently near a given limiting value.

[17] Hardy and Riesz, l.c., page 50.

THEOREM 20. *(First Inversion Formula) If* $\lambda_n < \omega < \lambda_{n+1}$ *and if the modified exponential series* $g(s) = \sum_1^\infty a_m \exp(-\lambda_m s) G(\lambda_m s)$ *is convergent for* $s = \sigma_0 + it_0$ *where* $\sigma_0 > \xi(\lambda_1)$ *(ξ being given by lemma 3), then for* $c > \sigma_0$ *we have*

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{g(s)}{s}\frac{\exp(\omega s)}{G(\lambda_n s)}\,ds = a_n + \sum_1^{n-1} \frac{a_m}{2\pi i}\int_{c-i\infty}^{c+i\infty} \exp[-(\lambda_m - \omega)s]\,\frac{G(\lambda_m s)}{G(\lambda_n s)}\frac{ds}{s}.$$

PROOF. We consider the function

$$(8.13)\qquad \begin{aligned} r(s) &= \frac{\exp[\omega(s-\sigma_0)]}{G(\lambda_n s)}\left\{g(s) - \sum_1^n a_m \exp(-\lambda_m s)G(\lambda_m s)\right\} \\ &= \sum_{m=n+1}^{\infty} a_m \exp[-(\lambda_m - \omega)s]\,\frac{G(\lambda_m s)}{G(\lambda_n s)}\exp(-\omega\sigma_0) \end{aligned}$$

where n is determined so that $\lambda_n < \omega < \lambda_{n+1}$.

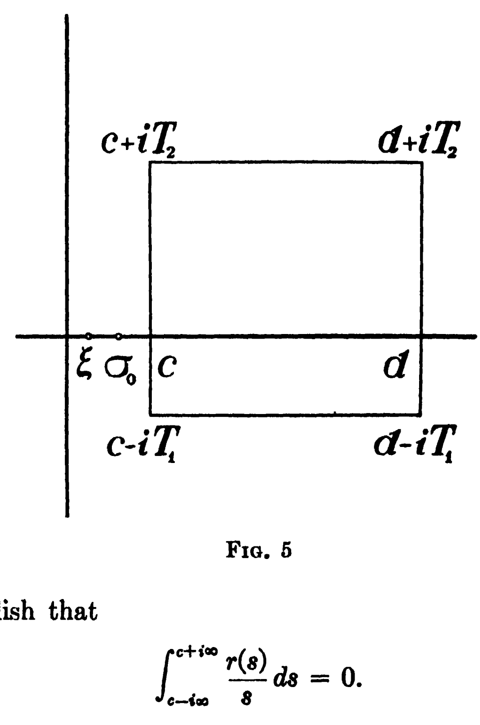


FIG. 5

We shall establish that

$$\int_{c-i\infty}^{c+i\infty} \frac{r(s)}{s}\,ds = 0.$$

We consider the closed rectangle as shown at the side. The function $r(s)/s$ has no singularities in this rectangle, and so by Cauchy's theorem we have

$$(8.14)\qquad \int_{c-iT_1}^{c+iT_2} \frac{r(s)}{s}\,ds = \int_{c-iT_1}^{d-iT_1} \frac{r(s)}{s}\,ds - \int_{c+iT_2}^{d+iT_2} \frac{r(s)}{s}\,ds + \int_{d-iT_1}^{d+iT_2} \frac{r(s)}{s}\,ds.$$

Now this finite rectangle can be enclosed in an angular sector about the point $s_0 = \sigma_0 + it_0$, and by using the same argument as we did in theorem 2, given

$\epsilon > 0$, there exists an integer n_0 such that $n_0 \geqq n$ and for which for $p \geqq n_0$ we have

$$\begin{aligned}(8.15)\qquad &\left|\sum_{m=p}^{\infty} a_m \exp\left[-(\lambda_m - \omega)s\right] \frac{G(\lambda_m s)}{G(\lambda_n s)} \exp(-\omega\sigma_0)\right| \\ &\qquad \leqq \frac{\epsilon K_{10}(\sigma_0)}{K_3(\xi)} \frac{|s - \sigma_0|}{\Re(s - \sigma_0)} \exp\left[-\Re(s - \sigma_0)\lambda_p\right] \exp(-\omega\sigma_0) \exp(\Re(s)\omega) \\ &\qquad \leqq \frac{\epsilon K_{10}}{K_3} \frac{|s - \sigma_0|}{\Re(s - \sigma_0)} \exp\left[-(\lambda_p - \omega)\Re(s - \sigma_0)\right] \\ &\qquad \leqq \epsilon K_{11},\end{aligned}$$

this estimate being independent of s as long as s is in the angular sector, and hence independent of the number d. Thus it follows that there exists a constant K_{12} such that

$$\left|\sum_{m=n+1}^{\infty} a_m \exp\left[-(\lambda_m - \omega)s\right] \frac{G(\lambda_m s)}{G(\lambda_n s)} \exp(-\omega\sigma_0)\right| < K_{12},$$

for all s in the angular sector.

Therefore

$$\begin{aligned}(8.16)\qquad \left|\int_{d-iT_1}^{d+iT_2} \frac{r(s)}{s}\, ds\right| &\leqq K_{12}\left|\int_{d-iT_1}^{d+iT_2} \frac{ds}{s}\right| = K_{12}\left|\log\frac{d + iT_2}{d - iT_1}\right| \\ &\leqq K_{12}\left|\log\frac{1 + \dfrac{iT_2}{d}}{1 - \dfrac{iT_1}{d}}\right|,\end{aligned}$$

and as $d \to \infty$ we have

$$(8.17)\qquad \left|\log\frac{1 + \dfrac{iT_2}{d}}{1 - \dfrac{iT_1}{d}}\right| \to 0.$$

We shall now show that

$$\int_{c-iT_1}^{d-iT_1} \frac{r(s)}{s}\, ds \quad \text{and} \quad \int_{c+iT_2}^{d+Ti_2} \frac{r(s)}{s}\, ds$$

are both convergent as $d \to \infty$, and thus we can write (8.14) as

$$(8.18)\qquad \int_{c-iT_1}^{c+iT_2} \frac{r(s)}{s}\, ds = \int_{c-iT_1}^{\infty-iT_1} \frac{r(s)}{s}\, ds - \int_{c+iT_2}^{\infty+iT_2} \frac{r(s)}{s}\, ds.$$

If we let the equation

$$(8.19)\qquad r(s) = \frac{\exp\left[-(\lambda_{n+1} - \omega)s\right]}{G(\lambda_n s)}\, p(s)$$

define the function $p(s)$ we have

$$
(8.20)\quad
\begin{aligned}
p(s) &= r(s) \exp [(\lambda_{n+1} - \omega)s]G(\lambda_n s) \\
&= \left[\sum_{m=n+1}^{\infty} a_m \exp [-(\lambda_m - \lambda_{n+1})s]G(\lambda_m s) \exp (-\omega\sigma_0) \right].
\end{aligned}
$$

Now $r(s)$ is convergent at the point $s = \sigma_0$. Therefore the series for the function $p(s)$ converges for $s = \sigma_0$. Identifying

$$b_p = a_{n+p},$$

$$\lambda_{n+p} = \mu_p,$$

we see that theorem 19 is applicable to the function

$$p(s) = \sum_{p=1}^{\infty} b_p \exp [-(\mu_p - \mu_1)s]G(\mu_p s) \exp (-\omega\sigma_0).$$

Therefore, given $\epsilon_1 > 0$, we can choose T such that $T > 0$ and

$$| p(s) | < \epsilon_1 T_2 \quad \text{for} \begin{cases} s = \sigma + iT_2 \\ \sigma \geqq c \\ T_2 > T \end{cases}.$$

Then

$$
(8.21)\quad
\begin{aligned}
\left| \int_{c+iT_2}^{d+iT_2} \frac{r(s)}{s} ds \right| &= \left| \int_{c+iT_2}^{d+iT_2} \frac{p(s)}{s} \frac{\exp [-(\lambda_{n+1} - \omega)s]}{G(\lambda_m s)} ds \right| \\
&\leqq \epsilon_1 T_2 \left| \int_{c+iT_2}^{d+iT_2} \frac{\exp [-(\lambda_{n+1} - \omega)s]}{sG(\lambda_m s)} ds \right| \\
&\leqq \frac{\epsilon_1 T_2}{K_3(\xi)} \frac{1}{\sqrt{c^2 + T_2^2}} \int_c^{\infty} \exp [-(\lambda_{n+1} - \omega)s]\, ds \\
&< \frac{\epsilon_1}{K_3} \frac{1}{\lambda_{n+1} - \omega},
\end{aligned}
$$

since $\dfrac{1}{|s|} \leqq \dfrac{1}{\sqrt{c^2 + T_2^2}}$ and $\dfrac{T_2}{\sqrt{c^2 + T_2^2}} < 1$. Thus we conclude that

$$(8.22)\quad \lim_{T_2 \to \infty} \left| \int_{c+iT_2}^{\infty+iT_2} \frac{r(s)}{s} ds \right| = 0,$$

and likewise

$$(8.23)\quad \lim_{T_1 \to \infty} \left| \int_{c-iT_1}^{\infty-iT_1} \frac{r(s)}{s} ds \right| = 0.$$

Therefore from (8.19) and the above we obtain the evaluation

$$\int_{c-i\infty}^{c+i\infty} \frac{r(s)}{s} ds = 0,$$

and recalling the definition of $r(s)$ (equation (8.13)) we have

$$\int_{c-i\infty}^{c+i\infty} \frac{g(s)}{s} \frac{\exp [\omega(s-\sigma_0)]}{G(\lambda_n s)} ds$$
$$= \int_{c-i\infty}^{c+i\infty} \left[\sum_{m=1}^{n} a_m \exp [-\lambda_m s] \frac{G(\lambda_m s)}{G(\lambda_n s)} \right] \exp [\omega(s-\sigma_0)] \frac{ds}{s}.$$

Using lemma 10 to evaluate the term in which a_n is a factor we obtain the first inversion formula,

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{g(s)}{s} \frac{\exp (\omega s)}{G(\lambda_n s)} ds = a_n + \sum_{1}^{n-1} \frac{a_m}{2\pi i} \int_{c-i\infty}^{c+i\infty} \exp [-(\lambda_m - \omega)s] \frac{G(\lambda_m s)}{G(\lambda_n s)} \frac{ds}{s}.$$

THEOREM 21. (*Second Inversion Formula*) *If the series*

$$g(s) = \sum_{m=1}^{\infty} a_m \exp (-\lambda_m s) G(\lambda_m s)$$

is convergent for $s = \sigma_0 + it_0$ *and if* $\lambda_n < \omega < \lambda_{n+1}$, *then for* $c \geqq \sigma_0$ *we have*

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{g(s) \exp (\omega s)}{s} ds = \frac{1}{2\pi i} \sum_{1}^{n} a_m \int_{c-i\infty}^{c+i\infty} \exp [-(\lambda_m - \omega)s] G(\lambda_m s) \frac{ds}{s}.$$

PROOF. The proof follows along the same lines as did the proof of theorem 20 except that we do not have to introduce the factor $1/G(\lambda_n s)$.

REMARKS: For the Hankel series and for the special cases when ν is an odd half-integer, the second inversion formula can be simplified. Let $\nu - \frac{1}{2} = k$ where k is an integer. It will be recalled that the Bessel functions have simple expressions for such values, and likewise the function $M_\nu(\lambda_m s)$ can be expressed by the finite series

$$\begin{aligned} M_\nu(\lambda_m s) &= \int_0^\infty \exp (-u) u^k \left(1 + \frac{u}{2\lambda_m s}\right)^k du \\ &= \int_0^\infty \exp (-u) u^k \left[1 + \binom{k}{1} \frac{u}{2\lambda_m s} + \binom{k}{2} \left(\frac{u}{2\lambda_m s}\right)^2 + \cdots + \left(\frac{u}{2\lambda_m s}\right)^k\right] du \\ &= k! + \binom{k}{1} \frac{(k+1)!}{2\lambda_m s} + \binom{k}{2} \frac{(k+2)!}{(2\lambda_m s)^2} + \cdots + \frac{(2k)!}{(2\lambda_m s)^k}, \end{aligned}$$

and on substituting this in theorem 21 and using lemma 10 we get

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{h(s) \exp (\omega s)}{s} ds = \sum_{1}^{n} a_m \left[k! + \binom{k}{1} \frac{(k+1)!}{2\lambda_m} \frac{(w - \lambda_m)}{1!} \right.$$
$$\left. + \binom{k}{2} \frac{(k+2)!}{(2\lambda_m)^2} \frac{(w-\lambda_m)^2}{2!} + \cdots + \frac{(2k)!}{(2\lambda_m)^k} \frac{(w-\lambda_m)^k}{k!} \right].$$

We recall that for $\nu = \frac{1}{2}$ the Hankel series reduces to the Dirichlet series, and we note that the above formula reduces to Perron's formula (8.1) for $\nu - \frac{1}{2} = k = 0$.

9. Conclusion

By defining

$$
(9.1)\qquad
\begin{aligned}
A(x) &= 0, \qquad 0 \leqq x < \lambda_1,\\
A(\lambda_1) &= \frac{a_1}{2},\\
A(\lambda_n) &= \sum_{m=1}^{n-1} a_m + \frac{a_n}{2}\\
A(x) &= \sum_{m=1}^{n} a_m, \qquad \lambda_n < x < \lambda_{n+1},
\end{aligned}
$$

we can express the Dirichlet series $f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ as the Stieltjes integral $f(s) = \int_0^{\infty} \exp(-xs)\, d(A(x))$, and Perron's formula for the sum of the coefficients becomes

$$
(9.2)\qquad A(x) = \frac{1}{2\pi i} \int_{-i\infty}^{c+i\infty} \frac{f(s)}{s} \exp(xs)\, ds.
$$

Likewise the modified exponential series $g(s) = \sum_{1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ becomes the Stieltjes integral

$$
(9.3)\qquad g(s) = \int_0^{\infty} \exp(-xs) G(xs)\, d(A(x)).
$$

It should be possible to obtain a convergence theorem similar to theorem 2 from a consideration of the Laplace-Stieltjes transform.[18]

An integration by parts of equation (9.3) would give formally

$$
(9.4)\qquad g(s) = \int_0^{\infty} s \exp(-xs) \left\{ G(xs) + \frac{\partial G(xs)}{\partial x} \right\} A(x)\, dx,
$$

which can be regarded as an integral equation with the kernel

$$
(9.5)\qquad s \exp(-xs) \left\{ G(xs) + \frac{\partial G(xs)}{\partial x} \right\},
$$

and the two inversion formulae for the modified exponential series could be written in the form

$$
(9.6)\qquad \int_0^{\infty} dA(x) \left[\int_{c-i\infty}^{c+i\infty} \exp[-(x-\omega)s] G(xs) \frac{ds}{s} \right] = \int_{c-i\infty}^{c+i\infty} \frac{g(s) \exp \omega s}{s}\, ds,
$$

and

[18] Private communication from Prof. J. D. Tamarkin to the author.

$$(9.7)\quad \int_0^{\infty} dA(x)\left[\int_{c-i\infty}^{c+i\infty} \exp\left[-(x-\omega)s\right]\frac{G(xs)}{G(\lambda_n s)}\frac{ds}{s}\right] = \int_{c-i\infty}^{c+i\infty}\frac{g(s)\exp \omega s}{sG(\lambda_n s)}\,ds,$$

as can be formally verified by substituting for $g(s)$, and changing the order of integration.

Hadamard's results relative to the composition of singularities of Taylor's series show that the only possible singular points for a function represented by the series $\sum_{n=1}^{\infty} a_n b_n \exp(-ns)$ are points of the form $\alpha + \beta$ where α is some singular point of the function $f(s) = \sum_{n=1}^{\infty} a_n \exp(-ns)$ and β is some singular point of $g(s) = \sum_{n=1}^{\infty} b_n \exp(-ns)$. Mandelbrojt obtained analogous results applicable to the general Dirichlet series, but the present author has been unable to get an extension of Mandelbrojt's work applicable to the modified exponential series.

In conclusion I wish to thank Prof. S. Bochner of Princeton University who suggested this topic and whose advice and criticism have always been helpful.

PRINCETON UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

ON THE INTEGRALS OF CANONICAL SYSTEMS

BY CARL LUDWIG SIEGEL

(Received December 19, 1940)

1. Trigonometrical series

We consider a *canonical system* of differential equations

$$\dot{x}_k = H_{y_k}, \qquad \dot{y}_k = -H_{x_k} \qquad (k = 1, \cdots, n) \tag{1}$$

and suppose that the real function H does not contain the independent variable t and is, in a neighbourhood of the origin, an *analytic* function of the $2n$ variables $x_1, \cdots, y_n$. Let the number n of degrees of freedom be at least 2. We suppose moreover that the origin is an *equilibrium point* of the system; i.e. all the $2n$ derivatives H_{x_k}, H_{y_k} $(k = 1, \cdots, n)$ vanish at the origin. It may also be assumed that the function H itself vanishes at the origin. Denoting the $2n$ variables $x_1, \cdots, y_n$ by $z_1, \cdots, z_{2n}$, we write the system (1) in the form

$$\dot{z}_k = \sum_{l=1}^{2n} a_{kl} z_l + R_k \qquad (k = 1, \cdots, 2n), \tag{2}$$

where R_k is a power series in $z_1, \cdots, z_{2n}$ beginning with terms of the second order.

Let $\lambda_1, \cdots, \lambda_{2n}$ be the characteristic roots of the matrix (a_{kl}). In our case of a canonical system, the characteristic polynomial is an even function; hence we may arrange the roots such that

$$\lambda_{k+n} = -\lambda_k \qquad (k = 1, \cdots, n). \tag{3}$$

We suppose that $\lambda_1, \cdots, \lambda_n$ are *linearly independent*, with respect to the field of rational numbers; this means that the relationship

$$g_1\lambda_1 + \cdots + g_n\lambda_n = 0$$

holds in integral numbers $g_1, \cdots, g_n$ only for $g_1 = 0, \cdots, g_n = 0$.

If w is a power series of $2n$ variables x_k, η_k $(k = 1, \cdots, n)$ without linear terms and the determinant $| w_{x_k \eta_l} | \neq 0$ at the origin, the equations

$$\xi_k = w_{\eta_k}, \qquad y_k = w_{x_k} \qquad (k = 1, \cdots, n) \tag{4}$$

define a *contact transformation* of the variables $x_1, \cdots, y_n$ into the variables $\xi_1, \cdots, \eta_n$, and the canonical system (1) is invariant:

$$\dot{\xi}_k = H_{\eta_k}, \qquad \dot{\eta}_k = -H_{\xi_k} \qquad (k = 1, \cdots, n). \tag{5}$$

It has been known[1] for a long time that after an appropriate contact transformation the function H will depend only upon the n products

$$\xi_k\eta_k = \zeta_k \qquad (k = 1, \cdots, n) \tag{6}$$

and take the form

$$H = \sum_{k=1}^{n} \lambda_k\zeta_k + R, \tag{7}$$

where R is a power series in the n variables $\zeta_1, \cdots, \zeta_n$ beginning with quadratic terms. The integration of (5) for this *normal form* of H is then immediate. Since

$$H_{\eta_k} = \xi_k H_{\zeta_k}, \qquad H_{\xi_k} = \eta_k H_{\zeta_k}, \tag{8}$$

we deduce from (5) that the n products ζ_k are constant; hence

$$\xi_k = \alpha_k e^{H_{\zeta_k} t}, \qquad \eta_k = \beta_k e^{-H_{\zeta_k} t} \qquad (k = 1, \cdots, n),$$

with arbitrary constants α_k, β_k and

$$\alpha_k\beta_k = \zeta_k \qquad (k = 1, \cdots, n).$$

The original unknown functions $x_1, \cdots, y_n$ become now series with the general term

$$c_{g_1\cdots g_n} e^{(g_1 H_{\zeta_1} + \cdots + g_n H_{\zeta_n})t},$$

where the coefficient $c_{g_1\cdots g_n}$ denotes a constant and $g_1, \cdots, g_n$ integers. If all the characteristic roots $\lambda_1, \cdots, \lambda_n$ are pure imaginary numbers and the initial values of $x_1, \cdots, y_n$ real, then all the values $H_{\zeta_1}, \cdots, H_{\zeta_n}$ are pure imaginary, and we get a representation of the solutions of the canonical system by trigonometrical series.

This elegant method of solution has also been generalized to the case of a function H which contains explicitly the variable t, in periodical form, and is closely related to the important researches of Delaunay, Hill and Poincaré[2] in celestial mechanics. However, there is a serious objection: *The question of convergence has never been settled.* If we define sum, difference, product, quotient and derivative of power series in a formal algebraic manner, we can perform these operations also with divergent power series, and then we can construct by straightforward calculation a transformation of the type (4) which reduces H to a power series of the n products $\xi_k\eta_k$ alone. But no proof for the convergence of this contact transformation has been given, with exception of some special examples, when the integration of the system (1) can also be carried out by elementary methods. On account of the small divisors appearing in the

[1] E. T. Whittaker, *On the solution of dynamical problems in terms of trigonometric series*, Proceedings of the London Mathematical Society, vol. 34 (1902), pp. 206-221. Cf. also G. D. Birkhoff, *Dynamical systems*, New York (1927), chap. 3, and E. T. Whittaker, *A treatise on the analytical dynamics of particles and rigid bodies*, 4th edition, Cambridge (1937), chap. 16.

[2] Cf. H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, Paris (1893), vol. 2.

coefficients of the transformation, it seemed to be probable[3] that the series will diverge in general, but no single example had hitherto been found. From Poincaré's well-known theorem[4] on the analytic integrals of canonical differential equations we can only infer that those series do not uniformly converge, if $\lambda_1, \cdots, \lambda_n$ are variable complex parameters, whereas this theorem cannot be applied to a fixed function H.

We arrange the coefficients of H in a certain order and denote them by $h_1, h_2, h_3, \cdots$. We assume that the power series H converges in a neighbourhood of the origin and that the characteristic roots $\lambda_1, \cdots, \lambda_n$ are pure imaginary linearly independent numbers. The corresponding systems $(h_1, h_2, h_3, \cdots)$ form the points of a space Σ. A point of Σ is called *singular*, if the transformation of H into the normal form (7) *cannot* be performed by a *convergent* contact transformation (4), and else *regular*.

THEOREM 1. *Let* $(c_1, c_2, c_3, \cdots)$ *be a point of* Σ *and* $\epsilon_1, \epsilon_2, \epsilon_3, \cdots$ *an arbitrary sequence of positive numbers. Then a singular point* $(h_1, h_2, h_3, \cdots)$ *of* Σ *exists in the domain*

$$c_k - \epsilon_k < h_k < c_k + \epsilon_k \qquad (k = 1, 2, 3, \cdots).$$

This theorem asserts that the singular points are *everywhere dense* in Σ. We shall reduce the proof to that of another theorem concerning the *integrals* of a canonical system. It would be important to obtain also some information about the distribution of the regular points of Σ, but this seems to be rather a difficult problem. We do not know e.g., if the regular points are also everywhere dense in Σ and if they constitute an open connected set of points. In particular, it would be interesting to decide, whether H is regular or singular in the special case of the restricted problem of three bodies, with respect to the equilibrium solutions of Lagrange. But this seems to be beyond the power of the known methods of analysis.

2. Integrals

If P is any convergent or divergent power series of the $2n$ variables $x_1, \cdots, y_n$, we define the Poisson bracket (P, H) by the power series

$$(P, H) = \sum_{k=1}^{n} (P_{x_k} H_{y_k} - P_{y_k} H_{x_k}).$$

[3] G. D. Birkhoff, *Surface transformations and their dynamical applications*, Acta Mathematica, vol. 43 (1922), pp. 1–119. Cf. on the other hand G. W. Hill, *Remarks on the progress of celestial mechanics since the middle of the century*, Bulletin of the American Mathematical Society, 2nd series, vol. 2 (1896), pp. 125–136, and E. T. Whittaker, *On the adelphic integral of the differential equations of dynamics*, Proceedings of the Royal Society of Edinburgh, vol. 37 (1918), pp. 95–116.

[4] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, Paris (1892), vol. 1, chap. 5.

We call P an *integral* of the canonical system (1), if the equation

$$(P, H) = 0$$

holds identically in the variables; in other words, P is an integral, if the relationship

$$\dot{P} = 0$$

follows from (1).

The expression (P, H) is invariant under any contact transformation. By (6) and (8), we obtain

$$(\zeta_k, H) = 0 \qquad (k = 1, \cdots, n).$$

Introducing into $\zeta_k = \xi_k \eta_k$ the original variables $x_1, \cdots, y_n$, we find n power series

$$\zeta_k = P_k(x_1, \cdots, y_n) \qquad (k = 1, \cdots, n) \tag{9}$$

which are integrals of (1). Since the functional determinant of $\xi_1, \cdots, \eta_n$ as functions of $x_1, \cdots, y_n$ does not vanish identically, these n power series are certainly independent one from another, i.e. there exists no power series of the variables P_k with constant coefficients not all zero, which vanishes identically in the variables $x_1, \cdots, y_n$.

Obviously the function H itself and more generally any convergent power series in the single variable H is a *convergent* integral. If H is regular, in the sense of our former definition, there will exist a convergent contact transformation reducing H to the normal form (7), hence the integrals (9) will then also converge. Moreover, by (4), we deduce easily that the integral P_k cannot be expressed as a power series in H alone. Therefore Theorem 1 is contained in

THEOREM 2. *Let $(c_1, c_2, c_3, \cdots)$ be a point of Σ and $\epsilon_1, \epsilon_2, \epsilon_3, \cdots$ an arbitrary sequence of positive numbers. There exists a point $(h_1, h_2, h_3, \cdots)$ in the domain*

$$c_k - \epsilon_k < h_k < c_k + \epsilon_k \qquad (k = 1, 2, 3, \cdots)$$

such that any convergent integral of the corresponding canonical system (1) *is a power series in the single variable H.*

The proof of this theorem depends upon several lemmata.

LEMMA 1. *Any integral of* (1) *can be represented as a power series in the n integrals $P_1, \cdots, P_n$.*

PROOF: It is obvious that the sum, the difference, the product of two integrals and more generally any power series in a finite number of integrals without constant terms is again an integral. Let $P(x_1, \cdots, y_n)$ be any integral of (1). By the contact transformation (4), this integral becomes a power series in the variables $\xi_1, \cdots, \eta_n$ with the general term

$$c_{\alpha_1 \cdots \beta_n} \prod_{k=1}^{n} (\xi_k^{\alpha_k} \eta_k^{\beta_k}),$$

where $c_{\alpha_1 \cdots \beta_n}$ denotes a constant. Consider now the n differences

$$\alpha_k - \beta_k = g_k \qquad (k = 1, \cdots, n).$$

The sum of all special terms with $g_1 = 0, \cdots, g_n = 0$ is a power series T in the n products $\xi_k \eta_k = \zeta_k = P_k(x_1, \cdots, y_n)$, and the expression

$$P - T = J$$

is again an integral. If J is not identically zero, take a term j of least degree $(\alpha_1 + \beta_1) + \cdots + (\alpha_n + \beta_n) = d$ and calculate the terms of degree d in the power series (J, H). Since

$$(J, H) = 0,$$

we obtain from (6) and (7) the relationship

$$\sum_{k=1}^{n} \lambda_k (\xi_k j_{\xi_k} - \eta_k j_{\eta_k}) = 0$$

or

$$\sum_{k=1}^{n} \lambda_k g_k = 0.$$

This is impossible, the numbers $\lambda_1, \cdots, \lambda_n$ being linearly independent and the integers g_k not all zero. Hence $J = 0$ and $P = T$ a power series in the integrals $P_1, \cdots, P_n$.

3. Linear transformation

For our further purposes it is practical to introduce new variables $u_1, \cdots, v_n$ by a special *linear* contact transformation, which reduces the *quadratic* terms of $H(x_1, \cdots, y_n)$ to the normal form $\lambda_1 u_1 v_1 + \cdots + \lambda_n u_n v_n$. Obviously we find such a transformation, if we replace in (4) the power series w by the sum of its quadratic terms. Let

$$(10) \qquad z_k = \sum_{l=1}^{n} (c_{kl} u_l + c_{k,l+n} v_l) \qquad (k = 1, \cdots, 2n)$$

be this transformation,

$$(11) \qquad \mathfrak{C} = (c_{kl})$$

its matrix and

$$(12) \qquad H = E(u_1, \cdots, v_n) = \sum_{k=1}^{n} \lambda_k u_k v_k + \cdots$$

the power series for H as a function of the new variables. By (3) and (12), the transformed canonical system is

$$(13) \quad \dot{u}_k = E_{v_k} = \lambda_k u_k + \cdots, \quad \dot{v}_k = -E_{u_k} = \lambda_{k+n} v_k + \cdots \quad (k = 1, \cdots, n).$$

Let $\mathfrak{A}$ be the matrix of the coefficients a_{kl} in (2) and $\mathfrak{D}$ the diagonal matrix with the diagonal elements $\lambda_1, \cdots, \lambda_{2n}$. By (2), (10), (11), (13), we deduce

$$\mathfrak{A}\mathfrak{C} = \mathfrak{C}\mathfrak{D}$$

or more explicitly

$$\mathfrak{A}\mathfrak{c}_k = \mathfrak{c}_k\lambda_k \qquad (k = 1, \cdots, 2n),$$

where $\mathfrak{c}_k$ denotes the k^{th} column of the matrix $\mathfrak{C}$. Since $\lambda_1, \cdots, \lambda_n$ are linearly independent, the $2n$ characteristic roots $\lambda_1, \cdots, \lambda_{2n}$ are certainly different one from another. Therefore the general solution of the linear equation

$$\mathfrak{A}\mathfrak{x} = \mathfrak{x}\lambda_k$$

is

$$\mathfrak{x} = \mathfrak{c}_k\rho$$

with an arbitrary scalar factor ρ. On the other hand, λ_k is pure imaginary, and hence by (3), the bar denoting the passage to conjugate complex numbers,

$$\mathfrak{A}\bar{\mathfrak{c}}_k = \bar{\mathfrak{c}}_k\lambda_{k+n} \qquad (k = 1, \cdots, n).$$

This proves the relationship

$$\bar{\mathfrak{c}}_k = \mathfrak{c}_{k+n}\rho_k \qquad (k = 1, \cdots, n); \tag{14}$$

with a certain scalar factor ρ_k. Hence the linear functions z_k of the variables $u_1, \cdots, v_n$ are not changed, if $\mathfrak{c}_l$ is replaced by $\bar{\mathfrak{c}}_l$ $(l = 1, \cdots, 2n)$ and at the same time the variables u_l, v_l by $\hat{u}_l, \hat{v}_l$, where

$$\hat{u}_l = \rho_l^{-1}v_l, \qquad \hat{v}_l = \bar{\rho}_l u_l \qquad (l = 1, \cdots, n); \tag{15}$$

For any power series F, we denote by $\bar{F}$ the power series with the conjugate complex coefficients and the same variables. Since H is a real function,

$$H(x_1, \cdots, y_n) = \bar{H}(x_1, \cdots, y_n),$$

we obtain

$$E(u_1, \cdots, v_n) = \bar{E}(\hat{u}_1, \cdots, \hat{v}_n)$$

and in particular, by (12) and (15),

$$\lambda_k = \bar{\lambda}_k\bar{\rho}_k\rho_k^{-1} \qquad (k = 1, \cdots, n).$$

The characteristic roots $\lambda_1, \cdots, \lambda_n$ are pure imaginary; hence the same holds for the n numbers $\rho_1, \cdots, \rho_n$. Obviously the canonical form of the system (13) and the quadratic terms of E are unchanged, if the variables u_k, v_k are replaced by $a_k^{-1}u_k, a_kv_k$ $(k = 1, \cdots, n)$, where $a_1, \cdots, a_n$ are arbitrary constants $\neq 0$. Denoting the columns $\mathfrak{c}_ka_k^{-1}$, $\mathfrak{c}_{k+n}a_k$ again by $\mathfrak{c}_k, \mathfrak{c}_{k+n}$, we have to replace the factor ρ_k by $\rho_k(a_k\bar{a}_k)^{-1}$. Therefore we may assume that $\rho_k = \pm i$.

If $\rho_k = +i$ for any value of k, we replace moreover u_k, v_k by v_k, $-u_k$ and obtain the case $\rho_k = -i$. Hence it is allowed to assume

$$\rho_k = -i \qquad (k = 1, \cdots, n).$$

Now we change the notation. We suppose that the function $H = H^* (x_1, \cdots, y_n)$ and the matrix $\mathfrak{C}$ are given and denote the characteristic roots $\lambda_1, \cdots, \lambda_n$ by $\lambda_1^*, \cdots, \lambda_n^*$, the corresponding power series E by E^*. Let P be the set of all convergent power series

$$E(u_1, \cdots, v_n) = \sum_{k=1}^{n} \lambda_k u_k v_k + \cdots \tag{16}$$

satisfying the condition

$$E(u_1, \cdots, v_n) = \bar{E}(iv_1, \cdots, iu_n),$$

with linearly independent values of $\lambda_1, \cdots, \lambda_n$. Then the inverse of the linear transformation (10) takes any such function E again into a real power series H of the variables $x_k = z_k$, $y_k = z_{k+n}$ $(k = 1, \cdots, n)$.

If F is any power series in the variables $u_1, \cdots, v_n$, we denote by F_l $(l = 0, 1, 2, \cdots)$ the sum of its terms of order l and by $\boxed{F_l}$ the maximum of the absolute values of the coefficients of these terms. The proof of Theorem 2 is now tantamount to the proof of

THEOREM 3. *Let $\epsilon_2, \epsilon_3, \epsilon_4, \cdots$ be an arbitrary sequence of positive numbers. There exists a power series E of* P *such that*

$$\boxed{E_l - E_l^*} < \epsilon_l \qquad (l = 2, 3, \cdots)$$

and any convergent integral of the canonical system

$$\dot{u}_k = E_{v_k}, \qquad \dot{v}_k = -E_{u_k} \qquad (k = 1, \cdots, n)$$

is a power series in the single variable E.

Let E be an arbitrary power series of the set P and $H(x_1, \cdots, y_n)$ the corresponding real function. By the linear transformation (10), the n power series

$$\zeta_k = P_k(x_1, \cdots, y_n) \qquad (k = 1, \cdots, n) \tag{17}$$

become integrals of the system

$$\dot{u}_k = E_{v_k}, \qquad \dot{v}_k = -E_{u_k} \qquad (k = 1, \cdots, n). \tag{18}$$

Let Q be one of these integrals and

$$Q_2 = \sum_{k,l=1}^{n} (\alpha_{kl} u_k u_l + \beta_{kl} u_k v_l + \gamma_{kl} v_k v_l)$$

the sum of its terms of second order, where

$$\alpha_{kl} = \alpha_{lk}, \qquad \gamma_{kl} = \gamma_{lk} \qquad (k, l = 1, \cdots, n).$$

Calculating the terms of second order on the left-hand side of the equation

$$(Q, E) = 0,$$

we find by (16) the expression

$$\sum_{k,l=1}^{n} \{\alpha_{kl}(\lambda_k + \lambda_l)u_k u_l + \beta_{kl}(\lambda_k - \lambda_l)u_k v_l - \gamma_{kl}(\lambda_k + \lambda_l)v_k v_l\}.$$

Hence

$$\alpha_{kl} = 0, \qquad \gamma_{kl} = 0 \qquad\qquad (k, l = 1, \cdots, n),$$

$$\beta_{kl} = 0 \qquad\qquad (k \neq l; k, l = 1, \cdots, n),$$

and ζ_k takes the special form

$$\zeta_k = \sum_{l=1}^{n} b_{kl} u_l v_l + \cdots \qquad\qquad (k = 1, \cdots, n). \tag{19}$$

By (4) and (10), the variables $\xi_1, \cdots, \eta_n$ can be expressed as power series in $u_1, \cdots, v_n$. The determinant of the linear terms in these power series is different from zero. Let $\psi_1, \cdots, \psi_n$ be indeterminates and

$$\chi_l = \sum_{k=1}^{n} b_{kl}\psi_k \qquad\qquad (l = 1, \cdots, n). \tag{20}$$

By (6), (19), (20), the quadratic form

$$\sum_{k=1}^{n} \psi_k \xi_k \eta_k$$

of the $2n$ variables $\xi_1, \cdots, \eta_n$ has the same rank as the quadratic form

$$\sum_{l=1}^{n} \chi_l u_l v_l$$

of the $2n$ variables $u_1, \cdots, v_n$. Hence the n linear forms (20) vanish simultaneously if and only if $\psi_k = 0$ $(k = 1, \cdots, n)$, and consequently their determinant is different from zero. On solving for $u_l v_l$ from (19) we obtain n integrals of (18) in the form

$$S^{(l)} = u_l v_l + \cdots \qquad\qquad (l = 1, \cdots, n).$$

Let us suppose that $S^{(l)}$ contains a term of the special type

$$c \prod_{k=1}^{n} (u_k v_k)^{\alpha_k}, \qquad \alpha_1 + \cdots + \alpha_n = g > 1 \tag{21}$$

and that the degree g of this term is as small as possible. Then the integral

$$S^{(l)} - c \prod_{k=1}^{n} S^{(k)\alpha_k}$$

does not contain a term of this special type of degree $\leqq g$. It is now obvious that we can also construct n integrals

$$s^{(l)} = u_l v_l + \cdots \qquad (l = 1, \cdots, n),$$

which do not contain *any* term of the form (21). Moreover the integrals $S^{(1)}, \cdots, S^{(n)}$ and consequently the integrals (17) are power series in $s^{(1)}, \cdots, s^{(n)}$. Lemma 1 gives the result, that any integral of (18) is a power series of the variables $s^{(1)}, \cdots, s^{(n)}$. This holds in particular for the integral E. By (16), the power series expressing E as a function of $s^{(1)}, \cdots, s^{(n)}$ has the form

$$E = \sum_{k=1}^{n} \lambda_k s^{(k)} + \cdots;$$

hence we may also represent $s^{(n)}$ as a power series in $E, s^{(1)}, \cdots, s^{(n-1)}$. Therefore we have

LEMMA 2. *There exist $n - 1$ integrals*

$$s^{(l)} = u_l v_l + \cdots \qquad (l = 1, \cdots, n - 1)$$

containing no term of the type (21) *and such that any integral of* (18) *is a power series in $E, s^{(1)}, \cdots, s^{(n-1)}$.*

For the rest of our investigation, we will only consider the case of two degrees of freedom. As a matter of fact, the generalization of our proof to the case $n > 2$ requires more complicated calculations, but does not present more serious difficulties.

4. Estimation of the coefficients

LEMMA 3. *Let ξ, η be complex numbers and G a homogeneous polynomial of degree r in the variables u_1, u_2, v_1, v_2. Then*

$$(22) \qquad (|\xi| + |\eta|)\lceil G \rceil \leqq (2r + 2)\lceil (\xi u_1 v_1 + \eta u_2 v_2)G \rceil.$$

PROOF: We can write

$$G = \sum_{l=0}^{r} G^{(l)},$$

where $G^{(l)}$ is also homogeneous in the variables u_1, u_2, of degree l. Since $\lceil G \rceil$ is the maximum of the values $\lceil G^{(l)} \rceil$ $(l = 0, \cdots, r)$ and $\lceil (\xi u_1 v_1 + \eta u_2 v_2)G \rceil$ the maximum of the values $\lceil (\xi u_1 v_1 + \eta u_2 v_2)G^{(l)} \rceil$, the inequality (22) will certainly hold, if it is true for $G^{(l)}$ instead of G and $l = 0, \cdots, r$. Hence we may suppose

$$G = \sum_{k=0}^{l} \varphi_k u_1^k u_2^{l-k},$$

where φ_k denotes a homogeneous polynomial in v_1, v_2, of degree $r - l$. If $|\xi| > |\eta|$, we interchange ξ, u_1, v_1 and η, u_2, v_2. Consequently, we have only to consider the case

$$(23) \qquad |\xi| \leqq |\eta|.$$

Putting

$$\varphi_{-1} = 0, \qquad \varphi_{l+1} = 0,$$

$$\Phi_k = \xi v_1 \varphi_{k-1} + \eta v_2 \varphi_k \qquad (k = 0, \cdots, l+1),$$

we obtain

$$(\xi u_1 v_1 + \eta u_2 v_2) G = \sum_{k=0}^{l+1} \Phi_k u_1^k u_2^{l+1-k}$$

$$(\eta v_2)^{k+1} \varphi_k = \sum_{p=0}^{k} (-\xi v_1)^{k-p} (\eta v_2)^p \Phi_p \qquad (k = 0, \cdots, l),$$

whence, by (23),

$$|\eta| \overline{|\varphi_k|} \leqq \sum_{p=0}^{k} \overline{|\Phi_p|}.$$

Since $\overline{|G|}$ is now the maximum of the numbers $\overline{|\varphi_k|}$ $(k = 0, \cdots, l)$ and $\overline{|(\xi u_1 v_1 + \eta u_2 v_2) G|}$ the maximum of the numbers $\overline{|\Phi_k|}$ $(k = 0, \cdots, l+1)$, the inequality

$$|\eta| \overline{|G|} \leqq (l+1) \overline{|(\xi u_1 v_1 + \eta u_2 v_2) G|}$$

holds. Moreover

$$|\xi| + |\eta| \leqq 2|\eta|, \qquad l + 1 \leqq r + 1,$$

and the lemma is proved.

By Lemma 2, the existence of $n - 1$ integrals $s^{(l)}$ $(l = 1, \cdots, n-1)$ with certain properties was stated. In our case $n = 2$, let us denote the integral $s^{(1)}$ more shortly by s. Then

$$s = \sum_{k=2}^{\infty} s_k, \qquad s_2 = u_1 v_1, \tag{24}$$

where s_k is a homogeneous polynomial in u_1, u_2, v_1, v_2, of degree k. If $k > 2$, the polynomial s_k does not contain a term of the special form $c(u_1 v_1)^\alpha (u_2 v_2)^\beta$.

Lemma 4. *Let the canonical system*

$$\dot{u}_k = E_{v_k}, \qquad \dot{v}_k = -E_{u_k} \qquad (k = 1, 2) \tag{25}$$

possess a convergent integral, which is not a power series in E alone. Then the sequence

$$\frac{\log \overline{|s_k|}}{k \log k} \qquad (k = 2, 3, \cdots)$$

has a finite upper bound.

Proof: By Lemma 2, any integral $P(u_1, u_2, v_1, v_2)$ of (25) can be written as a power series in E and s,

$$P = \sum_{\alpha, \beta = 0}^{\infty} c_{\alpha\beta} s^\alpha E^\beta.$$

We assume that there exists at least one coefficient

$$c_{\alpha\beta} \neq 0, \qquad \alpha > 0;$$

take $\alpha + \beta = g$ as small as possible. If P is a *convergent* integral, the same holds for the expression

$$p(u_1, u_2, v_1, v_2) = P - \sum_{\beta=0}^{g-1} c_{0\beta} E^{\beta}.$$

By (16) and (24), we find the decomposition

$$p = \sum_{k=2g}^{\infty} p_k,$$

where

$$p_{2g} = \sum_{\alpha+\beta=g} c_{\alpha\beta} s_2^{\alpha} E_2^{\beta}, \qquad s_2 = u_1 v_1, \qquad E_2 = \lambda_1 u_1 v_1 + \lambda_2 u_2 v_2.$$

The polynomial

$$\Delta = \frac{\partial p_{2g}}{\partial s_2} = \sum_{\alpha+\beta=g} \alpha c_{\alpha\beta} s_2^{\alpha-1} E_2^{\beta} \tag{26}$$

does not vanish identically; since it is homogeneous in the two variables $u_1 v_1$ and $u_2 v_2$, it may be written in the form

$$\Delta = c \prod_{h=1}^{g-1} (\xi^{(h)} u_1 v_1 + \eta^{(h)} u_2 v_2) \tag{27}$$

with constants $\xi^{(h)}$, $\eta^{(h)}$ not both zero ($h = 1, \cdots, g - 1$) and a constant $c \neq 0$.

We denote by x, y, z any three of the variables u_1, u_2, v_1, v_2. Since p is a power series in s and E, the functional determinant

$$\frac{d(E, p, s)}{d(x, y, z)} = 0,$$

identically in x, y, z. Calculating the terms of degree $h - 3$, we obtain

$$\sum_{\alpha+\beta+\gamma=h} \frac{d(E_\alpha, p_\beta, s_\gamma)}{d(x, y, z)} = 0 \qquad (h \geqq 2g + 4).$$

We apply this relationship for $x, y, z = u_1, u_2, v_1$ and for $x, y, z = u_1, u_2, v_2$. Denoting the corresponding expressions

$$\sum_{\alpha+\beta=l} \frac{d(E_\alpha, p_\beta)}{d(y, z)}, \qquad \sum_{\alpha+\beta=l} \frac{d(E_\alpha, p_\beta)}{d(z, x)}, \qquad \sum_{\alpha+\beta=l} \frac{d(E_\alpha, p_\beta)}{d(x, y)} \qquad (l \geqq 2g + 2)$$

by A_{1l}, A_{2l}, A_{3l} and B_{1l}, B_{2l}, B_{3l}, we find

$$\sum_{l+\gamma=h} \left(A_{1l} \frac{\partial s_\gamma}{\partial u_1} + A_{2l} \frac{\partial s_\gamma}{\partial u_2} + A_{3l} \frac{\partial s_\gamma}{\partial v_1} \right) = 0, \tag{28}$$

$$\sum_{l+\gamma=h} \left(B_{1l} \frac{\partial s_\gamma}{\partial u_1} + B_{2l} \frac{\partial s_\gamma}{\partial u_2} + B_{3l} \frac{\partial s_\gamma}{\partial v_2} \right) = 0. \tag{29}$$

Let $\mu_1, \mu_2, \mu_3, \mu_4$ be certain appropriate positive constants, which do not depend upon the subscript k appearing in the formulas. Since the power series E and p converge, the inequalities

$$\overline{|E_k|} < \mu_1^k \qquad (k \geqq 2),$$

$$\overline{|p_k|} < \mu_1^k \qquad (k \geqq 2g)$$

hold, and consequently

$$\overline{\left|\frac{\partial E_k}{\partial x}\right|} < k\mu_1^k, \qquad \overline{\left|\frac{\partial p_k}{\partial y}\right|} < k\mu_1^k.$$

The polynomial $\frac{\partial E_k}{\partial x}$ is a sum of $\binom{k+2}{3}$ terms, hence

$$\overline{\left|\frac{d(E_\alpha, p_\beta)}{d(x, y)}\right|} < 2\alpha\beta\binom{\alpha+2}{3}\mu_1^{\alpha+\beta}.$$

If ψ_k denotes any one of the six polynomials $A_{1k}, \cdots, B_{3k}$, we have

$$\overline{|\psi_k|} < \mu_2^k \qquad (k \geqq 2g+2),$$

whence

$$\overline{\left|\psi_l \frac{\partial s_\gamma}{\partial x}\right|} \leqq \gamma\binom{\gamma+2}{3}\mu_2^l\,\overline{|s_\gamma|}. \tag{30}$$

By (26), the identity

$$\frac{d(E_2, p_{2g})}{d(x, y)} = \Delta\frac{d(E_2, s_2)}{d(x, y)}$$

holds and therefore we obtain in the case $l = 2g + 2$ for $A_{1l}, \cdots, B_{3l}$ the values $\lambda_2 u_1 v_2 \Delta$, 0, $-\lambda_2 v_1 v_2 \Delta$, 0, $\lambda_2 v_1 u_2 \Delta$, $-\lambda_2 v_1 v_2 \Delta$. From (28), (29), (30), we deduce now the inequality

$$\overline{\left|\lambda_2 v_2 \Delta\left(u_1\frac{\partial s_k}{\partial u_1} - v_1\frac{\partial s_k}{\partial v_1}\right)\right|} + \overline{\left|\lambda_2 v_1 \Delta\left(u_2\frac{\partial s_k}{\partial u_2} - v_2\frac{\partial s_k}{\partial v_2}\right)\right|}$$
$$< \sum_{\gamma=2}^{k-1}(\gamma+3)^4\,\overline{|s_\gamma|}\,\mu_2^{k+2g+2-\gamma}$$

and, by (27) and Lemma 3,

$$\overline{\left|u_1\frac{\partial s_k}{\partial u_1} - v_1\frac{\partial s_k}{\partial v_1}\right|} + \overline{\left|u_2\frac{\partial s_k}{\partial u_2} - v_2\frac{\partial s_k}{\partial v_2}\right|} < k^{g+3}\sum_{\gamma=2}^{k-1}\overline{|s_\gamma|}\,\mu_3^{k-\gamma} \qquad (k \geqq 3). \tag{31}$$

If

$$\omega = a u_1^{\alpha_1} v_1^{\beta_1} u_2^{\alpha_2} v_2^{\beta_2}$$

is any term of s_k, then

$$|\alpha_1 - \beta_1| + |\alpha_2 - \beta_2| \geqq 1,$$

$$u_1 \frac{\partial\omega}{\partial u_1} - v_1 \frac{\partial\omega}{\partial v_1} = (\alpha_1 - \beta_1)\omega, \qquad u_2 \frac{\partial\omega}{\partial u_2} - v_2 \frac{\partial\omega}{\partial v_2} = (\alpha_2 - \beta_2)\omega,$$

and (31) implies

$$\overline{|s_k|} < k^{g+3} \sum_{\gamma=2}^{k-1} \overline{|s_\gamma|}\, \mu_3^{k-\gamma} \qquad (k = 3, 4, \cdots). \tag{32}$$

Obviously the inequality

$$\overline{|s_l|} \leqq (2l^{g+3}\mu_3)^{l-2} \tag{33}$$

is true for $l = 2$. If it is proved for $l = 2, 3, \cdots, k - 1$, we find from (32) the relationship

$$\overline{|s_k|} < k^{g+3} \sum_{\gamma=2}^{k-1} (2\gamma^{g+3}\mu_3)^{\gamma-2} \mu_3^{k-\gamma} \leqq (k^{g+3}\mu_3)^{k-2} \sum_{\gamma=2}^{k-1} 2^{\gamma-2} < (2k^{g+3}\mu_3)^{k-2},$$

and (33) holds for $l = k$. Hence

$$\overline{|s_k|} < k^{\mu_4 k}$$

$$\frac{\log \overline{|s_k|}}{k \log k} < \mu_4 .$$

5. Proof of Theorem 3

On account of Lemma 4, it is sufficient for the proof of Theorem 3 to construct a power series

$$E = \sum_{k=2}^{\infty} E_k$$

with the following 4 properties:

I) $$E_k(u_1, u_2, v_1, v_2) = \bar{E}_k(iv_1, iv_2, iu_1, iu_2) \quad (k = 2, 3, \cdots);$$

II) $$E_2 = \lambda_1 u_1 v_1 + \lambda_2 u_2 v_2,$$

where λ_1, λ_2 are linearly independent;

III) $$\overline{|E_k - E_k^*|} < \epsilon_k \qquad (k = 2, 3, \cdots);$$

IV) *the sequence*

$$\frac{\log \overline{|s_k|}}{k \log k} \qquad (k = 2, 3, \cdots)$$

has no finite upper bound.

The positive numbers $\epsilon_2, \epsilon_3, \cdots$ are arbitrarily given. Obviously we may suppose that

$$\epsilon_k < 1 \qquad (k = 2, 3, \cdots).$$

We begin with the construction of λ_1, λ_2. The coefficients λ_1^*, λ_2^* in

$$E_2^* = \lambda_1^* u_1 v_1 + \lambda_2^* u_2 v_2$$

are pure imaginary and linearly independent; hence

$$\omega^* = \frac{\lambda_1^*}{\lambda_2^*}$$

is a real number. We choose two integers q, r such that

$$q > 1 + |\omega^*| + 2|\lambda_2^*|\epsilon_2^{-1}, \tag{34}$$

$$|q\omega^* + r| < 1 \tag{35}$$

and define three sequences of numbers q_m, r_m, l_m $(m = 1, 2, \cdots)$ in the following manner:

$$r_m = q_m\left(\frac{r}{q} - \sum_{k=1}^{m} q_k^{-1}\right); \tag{36}$$

$$l_m = q_m + |r_m|; \tag{37}$$

$$q_1 = q^2; \tag{38}$$

q_{m+1} is the least integral power of q satisfying the inequality

$$q_{m+1} > q_m^2 + 4|\lambda_2^*|\epsilon_{l_m}^{-1} l_m^{m l_m}. \tag{39}$$

It is obvious that the numbers q_m, r_m, l_m are uniquely determined and q_m, l_m are positive. For the exponent a_m in

$$q_m = q^{a_m}$$

the inequality

$$a_{m+1} > 2a_m$$

holds. Since $a_1 = 2$, we obtain

$$a_m \geq 2^m, \qquad a_{m+1} - a_m > a_m \geq 2^m, \tag{40}$$

hence r_m and l_m are integers and the sequence $q_m q_{m+1}^{-1}$ tends to zero. By (34), (38), (40), the series

$$\theta = \sum_{k=1}^{\infty} q_k^{-1} \tag{41}$$

converges, and we find the estimations

$$0 < \sum_{k=h}^{\infty} q_k^{-1} < q_h^{-1} \sum_{k=0}^{\infty} q^{-k} \leqq 2q_h^{-1} \qquad (h = 1, 2, \cdots) \tag{42}$$

$$0 < \theta < 2q_1^{-1} \leqq q^{-1}. \tag{43}$$

Moreover, by (35), (36), (41),

$$| r_m | < q_m(| \omega^* | + q^{-1} + \theta),$$

whence, by (34), (37), (38), (39), (43),

$$l_{m+1} - l_m \geqq q_{m+1} - q_m - | r_m | > q_m(q^2 - 1 - | \omega^* | - 2q^{-1}) > 0. \tag{44}$$

Let λ_1, λ_2 be defined by

$$\lambda_1 = \lambda_2^* \left(\theta - \frac{r}{q}\right), \qquad \lambda_2 = \lambda_2^*; \tag{45}$$

then λ_1, λ_2 are pure imaginary and

$$E_2 = \lambda_1 u_1 v_1 + \lambda_2 u_2 v_2$$

satisfies I) with $k = 2$. For the expression

$$\rho_m = \lambda_1 q_m + \lambda_2 r_m \qquad (m = 1, 2, \cdots) \tag{46}$$

we find, by (36), (41), (45),

$$\rho_m = \lambda_2 q_m \sum_{k=m+1}^{\infty} q_k^{-1}$$

and by (42)

$$0 < \rho_m \lambda_2^{-1} < 2q_m q_{m+1}^{-1}. \tag{47}$$

If x, y are two indeterminates, the identity

$$(\lambda_1 x + \lambda_2 y) q_m \lambda_2^{-1} = (q_m y - r_m x) + \rho_m \lambda_2^{-1} x$$

holds. For given integral values of x and y, the number $q_m y - r_m x$ is integral, whereas the absolute value of $\rho_m \lambda_2^{-1} x$, by (47), is less than 1, if m is sufficiently large, and 0 only in the case $x = 0$. Therefore the numbers λ_1, λ_2 are linearly independent. On the other hand, by (34), (35), (43), (45),

$$| \lambda_1 - \lambda_1^* | = | \lambda_2^* | \left| \theta - \frac{r}{q} - \omega^* \right| < | \lambda_2^* | (q^{-1} + q^{-1}) < \epsilon_2.$$

Consequently II) is satisfied, and III) for $k = 2$.

By (34), (37), (38), we have $l_1 \geqq 4$. If we define

$$E_k = E_k^* \qquad (2 < k < l_1),$$

the conditions I) and III) are satisfied for $k < l_1$. Let m be any positive integer and assume, that the polynomials E_k are already determined for $k < l_m$ such that I) and III) hold. Consider now any power series $E = E_2 + E_3 + \cdots$ of P with these fixed terms E_k $(k = 2, 3, \cdots, l_m - 1)$. Let s be the integral

$$s = \sum_{k=2}^{\infty} s_k, \qquad s_2 = u_1 v_1$$

of the corresponding canonical system. Then

$$(s, E) = 0,$$

whence

$$\sum_{h=2}^{l} (s_h, E_{l+2-h}) = 0 \qquad (l = 3, 4, \cdots)$$

$$\lambda_1\left(u_1 \frac{\partial s_l}{\partial u_1} - v_1 \frac{\partial s_l}{\partial v_1}\right) + \lambda_2\left(u_2 \frac{\partial s_l}{\partial u_2} - v_2 \frac{\partial s_l}{\partial v_2}\right) = u_1 \frac{\partial E_l}{\partial u_1} - v_1 \frac{\partial E_l}{\partial v_1} - \sum_{h=3}^{l-1} (s_h, E_{l+2-h}). \tag{48}$$

Let $s_{\alpha\beta\gamma\delta} u_1^\alpha u_2^\beta v_1^\gamma v_2^\delta$, $E_{\alpha\beta\gamma\delta} u_1^\alpha u_2^\beta v_1^\gamma v_2^\delta$ be corresponding terms of s_l, E_l, where $s_{\alpha\beta\gamma\delta}$, $E_{\alpha\beta\gamma\delta}$ are the coefficients. Then $\alpha + \beta + \gamma + \delta = l$ and $s_{\alpha\beta\gamma\delta} = 0$ for $\alpha = \gamma$, $\beta = \delta$. From (48) we obtain

$$\{(\alpha - \gamma)\lambda_1 + (\beta - \delta)\lambda_2\} s_{\alpha\beta\gamma\delta} = (\alpha - \gamma) E_{\alpha\beta\gamma\delta} + b_{\alpha\beta\gamma\delta}, \tag{49}$$

where $b_{\alpha\beta\gamma\delta}$ denotes a certain bilinear function of the coefficients of $s_3, \cdots, s_{l-1}$ and $E_3, \cdots, E_{l-1}$. Since E_k is fixed for $k = 3, 4, \cdots, l_m - 1$, we infer from (49), that the coefficients of s_k for $k = 3, 4, \cdots, l_m - 1$ are uniquely determined and that the same holds for the expression

$$s_{\alpha\beta\gamma\delta} - \frac{\alpha - \gamma}{(\alpha - \gamma)\lambda_1 + (\beta - \delta)\lambda_2} E_{\alpha\beta\gamma\delta} \tag{50}$$

with $\alpha + \beta + \gamma + \delta = l_m$ and $\alpha - \gamma$, $\beta - \delta$ not both zero. Take in particular

$$\alpha = q_m, \quad \beta = r_m \quad \gamma = 0, \quad \delta = 0, \quad \text{if } r_m \geqq 0,$$
$$\alpha = q_m, \quad \beta = 0, \quad \gamma = 0, \quad \delta = -r_m, \quad \text{if } r_m < 0$$

and denote the corresponding coefficients of s_{l_m}, E_{l_m}, $\overset{*}{E}_{l_m}$ more shortly by σ_m, η_m, $\overset{*}{\eta}_m$. Then, by (46) and (50), the value

$$\sigma_m - q_m \rho_m^{-1}(\eta_m - \overset{*}{\eta}_m) = \nu_m \tag{51}$$

is uniquely given.

If we choose for η_m the two values

$$\eta_m = \overset{*}{\eta}_m \pm \tfrac{1}{2}\epsilon_{l_m}, \tag{52}$$

we find, by (51), two values of σ_m which have the difference $q_m\rho_m^{-1}\epsilon_{l_m}$. By (39), (45), (47), the inequality

$$|\, q_m\rho_m^{-1}\epsilon_{l_m}\,| > \tfrac{1}{2}q_{m+1}\,|\,\lambda_2\,|^{-1}\epsilon_{l_m} > 2l_m^{ml_m}$$

holds. Hence we can determine the sign in (52) such that

$$|\,\sigma_m\,| > l_m^{ml_m}. \tag{53}$$

By (44), we have $l_m < l_{m+1}$. For $k = l_m, l_m + 1, \cdots, l_{m+1} - 1$ we define E_k by

$$E_{l_m} = E^*_{l_m} \pm \tfrac{1}{2}\epsilon_{l_m}(u_1^{q_m}u_2^{r_m} + i^{l_m}v_1^{q_m}v_2^{r_m}), \qquad \text{if } r_m \geqq 0,$$

$$E_{l_m} = E^*_{l_m} \pm \tfrac{1}{2}\epsilon_{l_m}(u_1^{q_m}v_2^{-r_m} + i^{l_m}v_1^{q_m}u_2^{-r_m}), \qquad \text{if } r_m < 0,$$

$$E_k = E^*_k \qquad (l_m < k < l_{m+1}).$$

Now the conditions I) and III) are also satisfied for $l_m \leqq k < l_{m+1}$.

By this construction, a power series E satisfying I), II), III) is uniquely determined. The corresponding integral

$$s = \sum_{k=2}^{\infty} s_k$$

contains in the term s_{l_m} the coefficient σ_m. By (53), the inequality

$$\overline{|\, s_{l_m}|} > l_m^{ml_m} \qquad (m = 1, 2, \cdots)$$

holds. Hence the condition IV) is also satisfied, and the theorem is proved.

INSTITUTE FOR ADVANCED STUDY.

An extra number of the Annals of Mathematics will be published within a few months. This will be number 5 of volume 42. This extra number has been made possible by the generosity of the American Philosophical Society. The Index for the entire volume 42 will be incorporated in the extra number.

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

A PROPERTY OF INTEGRAL FUNCTIONS OF ORDER LESS THAN TWO WITH REAL ROOTS

BY K. S. K. IYENGAR

(Received March 29, 1940)

1. Introduction

P. Erdös and T. Grünwald in their paper "On polynomials with real roots," in *Annals of Mathematics*, Vol. 40, No. 3, July 1939, p. 537, have proved the following: Let $f(x)$ be a real polynomial with real roots having consecutive single roots at $x = a_1$, $x = a_2$, and let P be the point of intersection of the tangents to the curve $Y = f(x)$ at $(a_1, 0)$, $(a_2, 0)$, and let $M(a_1a_2)$ be maximum of $f(x)$ in (a_1a_2). Then

$$\tfrac{2}{3}M(a_1a_2)\,|a_2 - a_1| \geqq \left|\int_{a_1}^{a_2} f(x)\,dx\right| \geqq \tfrac{1}{3}\,|a_2 - a_1|\cdot|Y(P)| \tag{1.1}$$

where $Y(P)$ is the ordinate of the point P.

It is the object of this paper to give two very comprehensive theorems, and give the best possible inequality (3.3) of the type (1.1), from which it will be obvious that the inequality (1.1) is the best possible only in cases of symmetry.

2. Statement of the theorems

THEOREM 1: *Let $a_1 < a_2$, $m_1 > 0 > -m_2$, and let $\phi(x)$ be an integral function of order less than two, satisfying the following conditions.*

(1) *the roots of $\phi(x) = 0$ are all real*

(2) $\phi(a_1) = \phi(a_2) = 0$ *and* $\phi(x) \neq 0$ *in* $a_1 < x < a_2$

(3) $\phi'(a_1) = m_1$ *and* $\phi'(a_2) = -m_2$.

Then all the curves $Y = \phi(x)$ lie above the curve

$$Y = -A\cdot e^{B\cdot x}(x - a_1)(x - a_2) = m(x). \tag{2.1}$$

$$\textit{Where } A = (m_1m_2)^{1/2}/a_2 - a_1\cdot e^{-B/2(a_1+a_2)} \textit{ and } B = \frac{\log m_2/m_1}{a_2 - a_1}$$

and you can find polynomials with real roots satisfying conditions (2) *and* (3) *as near as you like to the minimum curve* (2.1).

THEOREM 2: *Let $a_1 < a_2$ and let $\phi(x)$ be an integral function of order < 2 such that,*

(1) *all roots of $\phi(x) = 0$ are real.*

(2) $\phi(a_1) = \phi(a_2) = 0$ *and* $\phi(x) > 0$ *in* $a_1 < x < a_2$ *and* $\phi'(a_1) \neq 0, \phi'(a_2) \neq 0$

(3) $\phi'(x_0) = 0, \quad a_1 < x_0 < a_2$

(4) $\phi(x_0) = M_0$.

Then all the curves $Y = \phi(x)$ *lie below the curve*

$$Y = \bar{A}(x - a_1)(a_2 - x)e^{\bar{B}x} = M(x) \tag{2.2}$$

$$\text{where } \bar{B} = \frac{2x_0 - a_1 - a_2}{(x_0 - a_1)(a_2 - x_0)} \quad \text{and} \quad \bar{A} = \frac{M_0 e^{-\bar{B}x_0}}{(x_0 - a_1)(a_2 - x_0)}$$

and you can find polynomials with real roots satisfying (2), (3) *and* (4) *as near as you like to the maximum curve* (2.2).

THEOREM 3_1: *The curve* $Y = Ae^{Bx}(x - a_1)(a_2 - x)$ *of theorem* 1 *as in* 2.1 *lies above the parabola which touches the lines* $Y = m_1(x - a_1)$ *and* $Y = -m_2(x - a_2)$ *at* $(a_1, 0)$ *and* $(a_2, 0)$ *and coincides with the minimum curve in case* $m_1 = m_2$.

THEOREM 3_2: *The curve* $Y = \bar{A}e^{\bar{B}x}(x - a_1)(a_2 - x)$ *of theorem* 2 *as in* 2.2 *lies below the parabola passing through* $(a_1, 0)$, $(a_2, 0)$ *and touching the line* $Y = M_0$ *at* x_0 *and if* $x_0 = \frac{a_1 + a_2}{2}$ *then the parabola will coincide with the maximum curve of* 2.2.

From theorems 1 and 3_1 we get

$$\begin{aligned} \int_{a_1}^{a_2} \phi(x)\,dx &\geqq \int_{a_1}^{a_2} m(x)\,dx \\ &= (a_2 - a_1)^2 \cdot (m_1 m_2)^{1/2} \cdot \left\{\frac{1}{6} + \frac{1}{2}\int_0^{B_1} \frac{u(\sinh u - u)}{B_1^3}\,du\right\} \end{aligned} \tag{3.1}$$

$\left[\text{where } B_1 = \frac{B(a_2 - a_1)}{2}, \; B \text{ being defined in 2.1}\right]$

$$\geqq \text{area of the parabola of contact} = \tfrac{1}{3}\, Y(P)(a_2 - a_1)^*$$

where $Y(P)$ is the Y-coordinate of the point of intersection of the tangents at $(a_1, 0)$, $(a_2, 0)$ to $Y = \phi(x)$.

From Theorems 2 and 3_2 we get

$$\int_{a_1}^{a_2} \phi(x)\,dx \leqq \int_{a_1}^{a_2} M(x)\,dx = \frac{e^{-2u_0^2/1-u_0^2}}{1 - u_0^2} \cdot 2M_0 \cdot (a_2 - a_1) \int_0^{B_2} \frac{u \cdot \sinh u}{B_2^3}\,du \tag{3.2}$$

$\left[\text{where } u_0 = x_0 - \frac{(a_1 + a_2)}{2} \Big/ \frac{a_2 - a_1}{2} \quad \text{and} \quad B_2 = 2u_0/1 - u_0^2\right]$

$$\leqq \text{area of the parabola in } (a_1 a_2) \text{ of theorem } 3_2 = \frac{2M_0}{3}(a_2 - a_1)$$[1]

---

[1] Note: The two elementary propositions, namely:

(i) if $Y = P(x)$ be a parabola touching the lines $Y = m_1(x - a_1)$
$Y = -m_2(x - a_2)$ at a_1, a_2 respectively

$$\int_{a_1}^{a_2} Y\,dx = \tfrac{1}{3}Y(P)(a_2 - a_1) = \frac{(a_2 - a_1)}{3} \cdot \left(\frac{m_1 m_2}{m_1 + m_2}\right)$$

so that combining (3.1) and (3.2) we get

$$\frac{2M_0}{3}(a_2 - a_1) \geqq \frac{e^{-2u_0^2/1-u_0^2}}{1-u_0^2} \cdot 2M_0(a_2 - a_1) \cdot \int_0^{B_2} \frac{u \cdot \sinh u}{B_2^3} du$$

$$\geqq \int_{a_1}^{a_2} \phi(x)\, dx \tag{3.3}$$

$$\geqq (a_2 - a_1)^2 (m_1 m_2)^{1/2} \left\{\frac{1}{6} + \frac{1}{2} \cdot \int_0^{B_1} \frac{u(\sinh u - u)}{B_1^3} du\right\}$$

$$\geqq \tfrac{1}{3} Y(P)(a_2 - a_1)$$

where $\phi(x)$ is any integral function of order < 2 satisfying the following conditions.

(1) $\phi(x) = 0$ are all real.

(2) $\phi(a_1) = \phi(a_2) = 0 \qquad \phi(x) > 0 \qquad$ in $a_1 < x < a_2$

(3) $\phi'(a_1) = m_1 \qquad \phi'(a_2) = -m_2$

(4) $\phi'(x_0) = 0 \qquad \phi(x_0) = M_0 \qquad a_1 < x_0 < a_2$.

It will be obvious in the course of the proofs of theorems 3_1 and 3_2 that equality in the first and last inequality signs in (3.3) can occur only in the case of symmetry, namely

$$m_1 = m_2 \quad \text{and} \quad x_0 = \frac{a_1 + a_2}{2}$$

so that the Erdös-Grünwald inequality is the best possible only in case of symmetry.

NOTE: We may generalize theorem 1 to the case where conditions (2) and (3) are generalized as follows:

(2) $\phi(a_1) = 0$ is a zero of pth degree $\qquad \phi(a_2) = 0$ of qth degree

(3) $\phi^p(a_1) = m_1 \qquad \phi^q(a_2) = m_2$.

Then the minimum curve will be of the type

$Y = A(x - a_1)^p (a_2 - x)^q e^{Bx}$ where A and B will be given in terms of m_1 and m_2.

A similar generalization for theorem 2 is also possible.

3. Proof of theorems

PROOF OF THEOREM 1. Let $F(x) = \dfrac{\phi(x)}{m(x)}$ where $m(x)$ is the function defined in (2.1); then

$$F(a_1) = F(a_2) = 1 \tag{4.1}$$

(ii) If $Y = P(x)$ be a parobola touching the line $Y = M_0$ at x_0 in $(a_1 a_2)$ and passing through $(a_1, 0)$, $(a_2, 0)$ then

$$\int_{a_1}^{a_2} y\, dx = \tfrac{2}{3} \cdot M_0(a_2 - a_1),$$

are assumed here.

and since $\phi(x)$ is an integral function of order < 2

$$\phi(x) = e^{\alpha x+\beta}(x - a_1)(a_2 - x)\left\{\prod_{n=3}^{\infty}\left(1 - \frac{x}{a_n}\right)e^{\frac{x}{a_n}}\right\} \tag{4.2}$$

all the a_n's for $n \geqq 3$ lying outside (a_1a_2). Hence

$$\frac{d^2}{dx^2}\log F(x) = -\sum_{n=3}^{\infty}\frac{1}{(x - a_n)^2} \tag{4.3}$$

so that $\log F(x)$ is concave (downwards) in the interval (a_1a_2) and hence the arc of $\log F(x)$ will lie above the chord in (a_1a_2) or

$$\log F(x) \geqq 0 \quad \text{or} \quad \phi(x) \geqq m(x) \qquad \text{in } (a_1a_2), \tag{4.4}$$

equality occurring only when $\phi(x)$ coincides with $m(x)$.

To prove the latter half of theorem 1, without any loss of generalization we take $a_1 = -1$, $a_2 = +1$. Let $m(x)$ be the corresponding minimum curve as defined in 2.1.

Let $P(x)$ be the polynomial $P(x) = (1 - x^2)\left(1 - \dfrac{x}{a}\right)^n$ a and n to be chosen presently, $\{\, |\, a\, | > 1\}$; n being given by $a = \dfrac{\left(\frac{m_2}{m_1}\right)^{\frac{1}{n}} + 1}{\left(\frac{m_2}{m_1}\right)^{\frac{1}{n}} - 1}$ so that $P'(-1) = m_1$, $P'(1) = -m_2$.

Let $m_2 > m_1$ (the argument being the same also when $m_2 < m_1$). Let

$$\theta(x) = \log\frac{P(x)}{m(x)},$$

then
$$\theta(\pm 1) = 0 \quad \text{and} \quad \theta''(x) = -\frac{n}{(x - a)^2}.$$

If $-1 < x < 1$

$$|\,\theta(x)\,| \leqq \frac{(1 - x^2)}{2}\cdot[\text{Max of } |\,\theta''(x)\,| \text{ in } (-1, 1)].$$

Now since $m_2 > m_1$, $(a > 1)$ and max. of $|\,\theta''\,|$ in $(-1, 1)$ will be at $x = 1$

$$|\,\theta''(1)\,| = \frac{n}{(a - 1)^2} = \frac{n}{4}\cdot\left\{\left(\frac{m_2}{m_1}\right)^{\frac{1}{n}} - 1\right\}^2$$

and for large n the last term will be $O\left(\dfrac{1}{n}\right)$, hence $P(x) = m(x)e^{O\left(\frac{1}{n}\right)} = m(x)(1 + \epsilon_n)$, thus proving theorem 1 completely.

Proof of Theorem 2: Let $F(x) = \log\dfrac{\phi(x)}{M(x)}$ where $M(x)$ is the maximum

function as defined in (2.2). Then

$$F(x_0) = F'(x_0) = 0,$$

and since ϕ is an integral function of order < 2,

$$\phi(x) = e^{\alpha x+\beta}(x - a_1)(a_2 - x)\cdot\left\{\prod_3^\infty \left(1 - \frac{x}{a_n}\right) e^{\frac{x}{a_n}}\right\}$$

a_n for $n \geqq 3$ lying outside (a_1a_2).

Since $$\frac{d^2}{dx^2} F(x) = -\sum_3^\infty \frac{1}{(x - a_n)^2} \leqq 0, \qquad \text{in } (a_1 a_2)$$

and $F'(x_0) = 0$ at x_0 in $(a_1 < x_0 < a_2)$ we see that $F(x) \leqq F(x_0) = 0$ in $(a_1 \leqq x \leqq a_2)$, hence $\phi(x) \leqq M(x)$.

The latter part of theorem 2, that we can find polynomials satisfying given condition as near as you like to $M(x)$ can be proved in the same manner as the latter part of theorem 1.

Proof of Theorem 3_1: Without loss of generality we may take $a_1 = -1$ $a_2 = 1$; given $m_2 \neq m_1$, the parabola of contact will be given by

$$Y = \sqrt{\alpha^2 + \beta^2 + 2\alpha\cdot\beta\cdot x} - \alpha - \beta x = P(x)$$

where $$\alpha = \frac{2m_1m_2(m_1 + m_2)}{(m_1 - m_2)^2} \qquad \beta = \frac{2m_1m_2}{(m_1 - m_2)}$$

Let $F(x) = \log \dfrac{m(x)}{P(x)}$ where $m(x)$ is the minimum curve as defined in (2.1) for the case $a_1 = -1$, $a_2 = 1$

Then

$$F(x) \equiv \log\left\{\frac{m(x)}{\beta^2(1 - x^2)}(\sqrt{\alpha^2 + \beta^2 + 2\alpha\beta x} + \alpha + \beta x)\right\}$$

$$F(\pm 1) = 0 \quad \text{and} \quad F''(x) = \frac{d^2}{dx^2}\log\theta(x),$$

where

$$\theta(x) = \sqrt{\alpha^2 + \beta^2 + 2\alpha\beta x} + \alpha + \beta x = \frac{\theta''}{\theta} - \frac{\theta'^2}{\theta^2};$$

since

$$\theta'' = \frac{-(\alpha\beta)^2}{(\alpha^2 + \beta^2 + 2\alpha\beta x)^{3/2}} \quad \text{and} \quad \theta(x) > 0 \text{ in } (-1, 1)$$

$$\frac{d^2}{dx^2}\log\theta(x) < 0 \text{ in } (-1, 1).$$

Hence F is concave downwards and the arc in $(-1, 1)$ lies above the chord, or

$$F(x) \geqq F(\pm 1) = 0$$

or $$m(x) \geqq P(x).$$

It is clear that in case $m_2 = m_1$ then the parabola of contact will be given by $Y = \frac{m_2}{2}(1 - x^2)$ and coincides with the minimum curve.

Proof of Theorem 3_2: As in theorem 3_1, $a_1 = -1$, $a_2 = 1$ and x_0 the maximum pt. $\neq 0$. Then the parabola

$$Y = \sqrt{\alpha^2 + \beta^2 + 2\alpha\beta x} - \alpha - \beta x$$

where

$$\alpha = \frac{M_0}{2x_0^2} \qquad \beta = -\frac{M_0}{x_0}$$

passes through $(-1, 0)$, $(1, 0)$ and touches $Y = M_0$ at (x_0, M).
Let $M(x)$ be the maximum function defined as in 2.2 and let

$$F(x) = \log \frac{M(x)}{P(x)}.$$

Then $$F(x_0) = F'(x_0) = 0$$

and arguing as in the previous paragraph

$$F''(x) \leqq 0 \qquad \text{in } (a_1 a_2).$$

Hence in $(a_1 a_2)$ $$F(x) \leqq F(x_0) = 0$$

or $$M(x) \leqq P(x)$$

and in case $x_0 = 0$ the parabola will be $Y = M(1 - x^2)$ and coincides with the maximum curve.

University of Mysore
Bangalore, India

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

THE PARTIAL SUMS OF MULTIVALENTLY STAR-LIKE FUNCTIONS

BY M. S. ROBERTSON

(Received September 12, 1940)

1. Introduction

In a recent paper (1) the author has a theorem concerning the n^{th} partial sum of an analytic function univalently star-like[1] with respect to the unit circle. The theorem 7.2 stated that the n^{th} partial sum was also univalently star-like with respect to the circle of radius $1 - 2n^{-1} \log n$. Dr. Otto Szász has kindly pointed out to the author that in the proof of this result an incorrect estimate for $| \rho_n'(z) |$, line 11, page 406, was used. The correct estimate should read

$$(1.1) \qquad | \rho_n'(z) | \leqq r^n(1 - r)^{-3} \cdot \{n^2(1 - r)^2 + (2n - 1)(1 - r) + 2\}$$

whence line 12 becomes

$$(1.2) \quad \Re\left[\frac{zf_n'(z)}{f_n(z)}\right] \geqq \frac{1 - r}{1 + r} - \frac{r^n(1 + r)^2}{1 - r}\left[\frac{\{(n - 1)(1 - r) + 2\}\{n(1 - r) + 2\}}{(1 - r)^2 - r^n(1 + r)^2(1 + n - nr)}\right]$$

which may be shown to be positive for $n > n_0$ when

$$(1.3) \qquad r = 1 - 4n^{-1} \log n$$

if one proceeds by the method used in the proof of Theorem 7.2. Thus Theorem 7.2 of the paper (1) should be corrected to read

THEOREM 7.2: *If*

$$f(z) = z + \sum_2^{\infty} a_n z^n$$

be regular and univalently star-like with respect to the unit circle then the n^{th} partial sum is univalently star-like for $| z | \leqq 1 - 4n^{-1} \log n$, $n > n_0$.

Whether the constant 4 is the best possible one or not the author is unable to say. There are reasons for belief that the best possible constant is the number 3. For instance, as will be shown in the last section of this paper, the well-known extremal function $z(1 - z)^{-2}$ of the theory of univalently star-like functions has a star radius R_n for its n^{th} partial sum with

$$\limsup_{n \to \infty} \left\{\frac{1 - R_n}{n^{-1} \log n}\right\} = 3.$$

Since Theorems 7.3, 7.4, and 7.5 of paper (1) were made to depend upon the result of Theorem 7.2, the expression $1 - 2n^{-1} \log n$ in the statements of each of

[1] Multivalently star-like functions of order p are defined in the lines immediately following equation (1.4) of the present paper; univalently star-like functions are those of order 1.

these theorems should be replaced by $1 - 4n^{-1} \log n$. Dr. Szász has informed the author that the sharp estimate $1 - 3n^{-1} \log n$ holds for Theorems 7.4 and 7.5. Instead of making Theorems 7.4 and 7.5 depend upon Theorem 7.2 Dr. Szász obtained his result by showing directly that the n^{th} partial sum of the function $z(1 - 2z \cos \theta + z^2)^{-1}$ is typically-real for $|z| \leqq 1 - 3n^{-1} \log n$. The above function, as was shown in Theorem 7.4, sets the pace for all the functions of the class under consideration. It might be pointed out that the function $z(1 - 2z \cos \theta + z^2)^{-1}$ has its n^{th} partial sum star-like univalently, and à fortiori typically real, for $|z| \leqq 1 - 3n^{-1} \log n$ as may be seen by direct calculation of $\Re[zf_n'(z)/f_n(z)]$ for the n^{th} partial sum $f_n(z)$ of this function. I omit the details but the proof is quite similar to that given for the function $z(1 - z)^{-2}$ in the last section of this paper.

It is now the purpose of this paper, besides correcting the error mentioned above, to generalize Theorem 7.2 to multivalently star-like functions. A definition of multivalently star-like functions follows.

Let

$$f(z) = z^p + a_{p+1}z^{p+1} + \cdots + a_n z^n + \cdots \tag{1.4}$$

be regular and multivalently star-like of order p with respect to the unit circle. This means that within the unit circle $f(z)$ assumes no value more than p times, at least one value p times, and in addition is star-like, i.e.,

$$\Re[zf'(z)/f(z)] > 0, \qquad |z| < 1. \tag{1.5}$$

Geometrically, this means that as z traverses a circle $|z| = r < 1$ in the anti-clockwise direction the radius vector joining the origin to the point $w = f(z)$ in the w-plane turns also in the anti-clockwise direction, completing in this case p revolutions as z traverses the circle once. When $p = 1$ $f(z)$ is said to be univalently star-like. In this case the radius vector cuts the image of $|z| = r < 1$ in the w-plane once only for any direction from the origin.

We have the following generalization of Theorem 7.2 mentioned above.

THEOREM A: *If*

$$f(z) = z^p + a_{p+1}z^{p+1} + \cdots + a_n z^n + \cdots$$

be regular and multivalently star-like of order p with respect to the unit circle then the n^{th} partial sum

$$f_n(z) = z^p + a_{p+1}z^{p+1} + \cdots + a_n z^n$$

for $n > n_0(p)$ is also multivalently star-like of order p for

$$|z| < 1 - (2p + 2)n^{-1} \log n.$$

I do not know if the constant $2p + 2$ can be replaced by a smaller one. However, the constant cannot be replaced by one which is smaller than $2p + 1$. For, as will be shown in the last section of the present paper, the multivalently

star-like function of order p $z^p(1-z)^{-2p}$ has its n^{th} partial sum multivalently star-like of order p in a circle about the origin whose radius is at most R_n where

$$\begin{aligned} &R_n \geqq 1 - (2p+1)n^{-1}\log n \\ &\limsup_{n\to\infty} \left\{\frac{1-R_n}{n^{-1}\log n}\right\} = 2p+1. \end{aligned} \tag{1.6}$$

In the proofs to follow A will denote a constant depending only upon p and the A in one inequality need not necessarily be the same A which occurs again in another or in the same inequality.

2. Preliminary lemmas

Before coming to the main theorem we prove four lemmas.

LEMMA 1. *If*

$$f(z) = z^p + a_{p+1}z^{p+1} + \cdots + a_nz^n + \cdots \tag{2.1}$$

be regular and multivalently star-like of order p for $|z| < 1$, *then*

$$f(z) = [\phi(z)]^p \tag{2.2}$$

where

$$\phi(z) = z + d_2z^2 + \cdots + d_nz^n + \cdots \tag{2.3}$$

is regular and univalently star-like for $|z| < 1$.

PROOF: Since $f(z)$ is star-like for $|z| < 1$ we have

$$\Re[zf'(z)/f(z)] > 0, \ |z| < 1. \tag{2.4}$$

Hence

$$p + z\frac{d}{dz}\log\{z^{-p}\cdot f(z)\} = zf'(z)/f(z) = pF(z) \tag{2.5}$$

where $F(z)$ is regular, $F(0) = 1$, $\Re F(z) > 0$ for $|z| < 1$. Thus

$$\begin{aligned} f(z) &= z^p\cdot\exp p\int_0^z z^{-1}\{F(z)-1\}\,dz \\ &= [\phi(z)]^p \end{aligned} \tag{2.6}$$

where

$$\begin{aligned} &\phi(z) = z\exp\int_0^z z^{-1}\{F(z)-1\}\,dz, \qquad \phi'(0) = 1, \\ &\Re[z\phi'(z)/\phi(z)] = \Re F(z) > 0. \end{aligned} \tag{2.7}$$

Thus $\phi(z)$ is univalently star-like and the proof is complete.

LEMMA 2. *If* $f(z)$ *satisfies the conditions of Lemma 1 then*

$$|f(z)| \geqq r^p(1+r)^{-2p}, \qquad |z| = r < 1, \tag{2.8}$$

$$\Re[zf'(z)/f(z)] \geqq p(1-r)(1+r)^{-1}, \tag{2.9}$$

$$|a_n| \leqq \frac{(n+p-1)!}{(n-p)!(2p-1)!}, \qquad n > p. \tag{2.10}$$

PROOF: We have from Lemma 1

$$f(z) = [\phi(z)]^p \tag{2.11}$$

and since $\phi(z)$ is univalent (2) for $|z| < 1$

$$|\phi(z)| \geqq r(1+r)^{-2}, \qquad |z| = r < 1. \tag{2.12}$$

Thus

$$|f(z)| \geqq r^p(1+r)^{-2p}. \tag{2.13}$$

Again, since

$$zf'(z)/f(z) = p \cdot F(z) \tag{2.14}$$

where $\Re F(z) > 0$, $F(0) = 1$, then (3)

$$\Re[zf'(z)/f(z)] = p\Re F(z) \geqq p(1-r)(1+r)^{-1}. \tag{2.15}$$

Finally, it is well known that $\phi(z)$ is majorized (2) by $z(1-z)^{-2}$ since $\phi(z)$ is star-like univalently. Hence $f(z)$ is majorized by $z^p(1-z)^{-2p}$ and thus

$$|a_n| \leqq \frac{(n+p-1)!}{(n-p)!\,(2p-1)!}, \qquad n > p. \tag{2.16}$$

This completes the proof of Lemma 2.

LEMMA 3. *The following identity holds for $r < 1$, p a positive integer.*

$$\sum_{\kappa=n+1}^{\infty} \kappa^p r^\kappa = \frac{r^{n+1}}{(1-r)^{p+1}}\left[p! + \sum_{m=1}^{p} A_m^p(1-r)^m\right], \qquad A_p^p = n^p, \tag{2.17}$$

where A_m^p = a polynomial in n of degree m whose leading coefficient is positive.

PROOF: It is easily verified that

$$\begin{aligned} \sum_{n+1}^{\infty} \kappa r^\kappa &= \frac{r^{n+1}}{(1-r)^2}[1 + n(1-r)], \\ \sum_{n+1}^{\infty} \kappa^2 r^\kappa &= \frac{r^{n+1}}{(1-r)^3}[2 + (2n-1)(1-r) + n^2(1-r)^2], \\ \sum_{n+1}^{\infty} \kappa^3 r^\kappa &= \frac{r^{n+1}}{(1-r)^4}[6 + (6n-6)(1-r) \\ &\qquad + (3n^2-3n+1)(1-r)^2 + n^3(1-r)^3]. \end{aligned} \tag{2.18}$$

Thus the identity (2.17) is true for $p = 1, 2, 3$. We prove that it is true for any positive integer p by mathematical induction. Assuming (2.17) true for $p = q$ we shall show that in this case it must also be true for $p = q + 1$. By assumption then (2.17) holds with $p = q$. Then if $A_0^q = q!$

$$
\begin{aligned}
\sum_{\kappa=n+1}^{\infty} \kappa^{q+1} r^{\kappa} &= r \frac{d}{dr} \sum_{m=0}^{q} A_m^q r^{n+1} (1-r)^{m-q-1} \\
&= r^{n+1}(1-r)^{-q-2} \cdot \sum_{m=0}^{q} A_m^q [(n+m-q) \\
&\qquad \times (1-r)^{m+1} - (m-q-1)(1-r)^m] \\
&= \frac{r^{n+1}}{(1-r)^{q+2}} \Big[(q+1) A_0^q + \sum_{m=0}^{q} \{ n A_m^q \\
&\qquad + (q-m)(A_{m+1}^q - A_m^q) \} (1-r)^{m+1} \Big].
\end{aligned} \tag{2.19}
$$

In the last summation the coefficient of $(1 - r)^{q+1}$ is seen to be n^{q+1} and the coefficient of $(1 - r)^{m+1}$ is of degree $(m + 1)$ in n since $q - m \geqq 0$ and the leading coefficient of A_m^q is positive for all m. Thus (2.17) holds for $p = q + 1$ whenever it is true for $p = q$. But since we know that it holds for $p = 1, 2, 3$ the proof by mathematical induction is complete.

Letting

$$
\begin{aligned}
f_n(z) &= z^p + a_{p+1} z^{p+1} + \cdots + a_n z^n, \qquad n \geqq p, \\
\rho_n(z) &= a_{n+1} z^{n+1} + a_{n+2} z^{n+2} + \cdots, \\
f(z) &= f_n(z) + \rho_n(z),
\end{aligned} \tag{2.20}
$$

we proceed to prove

LEMMA 4. *If* $|z| = r = 1 - (2p + 2)n^{-1} \log n$ *then for* $n > n_0(p)$ *there is a constant* A *depending upon* p *but not upon* n *for which*

$$
\begin{aligned}
|\rho_n(z)| &< A n^{-2} (\log n)^{-1}, \\
|\rho'_n(z)| &< A n^{-1} (\log n)^{-1}, \\
|f'(z)/f(z)| &< A n (\log n)^{-1}, \\
|f(z)| - |\rho_n(z)| &\geqq A.
\end{aligned} \tag{2.21}
$$

PROOF:

$$
|\rho_n(z)| \leqq \sum_{\kappa=n+1}^{\infty} \frac{(\kappa + p - 1)!}{(\kappa - p)!(2p-1)!} r^{\kappa} < A \sum_{n+1}^{\infty} \kappa^{2p-1} r^{\kappa}
$$

$$
< A r^{n+1} (1-r)^{-2p} [(2p-1)! + A_1^{2p-1}(1-r) + \cdots + n^{2p-1}(1-r)^{2p-1}]
$$

$$
r^{n+1} < \{1 - (2p+2)n^{-1} \log n\}^n < e^{-(2p+2)\log n} = n^{-2p-2}, \tag{2.22}
$$

$$
(2p-1)! + A_1^{2p-1}(1-r) + \cdots + n^{2p-1}(1-r)^{2p-1} < A(\log n)^{2p-1},
$$

$$
|\rho_n(z)| < A n^{-2p-2} \left\{ \frac{n}{(2p+2)\log n} \right\}^{2p} \cdot (\log n)^{2p-1} < A n^{-2} (\log n)^{-1}.
$$

Similarly we have

$$
(2.23)\quad
\begin{aligned}
|\rho_n'(z)| &< A\sum_{k=n}^{\infty} k^{2p}r^k \\
&< \frac{Ar^n}{(1-r)^{2p+1}}\,[(2p)! + A_1^{2p}(1-r) + \cdots + (n+1)^{2p}(1-r)^{2p}] \\
&< An^{-2p-2}\left\{\frac{n}{(2p+2)\log n}\right\}^{2p+1}\cdot(\log n)^{2p} < A(n\log n)^{-1}.
\end{aligned}
$$

Again,

$$
(2.24)\quad \left|\frac{f'(z)}{f(z)}\right| = \frac{p}{r}\,|F(z)| \leqq \frac{p}{r}\left(\frac{1+r}{1-r}\right) < An(\log n)^{-1},
$$

and

$$
(2.25)\quad |f(z)| - |\rho_n(z)| \geqq r^p(1+r)^{-2p} - An^{-2}(\log n)^{-1} > A,
$$

for $n > n_0(p)$, $r = 1 - (2p+2)n^{-1}\log n$.

3. Proof of Theorem A

Since

$$
(3.1)\quad f_n(z) = f(z) - \rho_n(z)
$$

we have for $r = 1 - (2p+2)n^{-1}\log n$

$$
(3.2)\quad
\begin{aligned}
\Re[zf_n'(z)/f_n(z)] &= \Re\left\{[zf'(z)/f(z)] - z\cdot\frac{\rho_n'(z) - \rho_n(z)\cdot f'(z)/f(z)}{f(z)-\rho_n(z)}\right\} \\
&\geqq \Re[zf'(z)/f(z)] - r\cdot\frac{|\rho_n'(z)| + |\rho_n(z)|\cdot|f'(z)/f(z)|}{|f(z)| - |\rho_n(z)|} \\
&\geqq p(1-r)(1+r)^{-1} \\
&\qquad - \frac{A(n\log n)^{-1} + An^{-2}(\log n)^{-1}\cdot n(\log n)^{-1}}{A} \\
&\geqq An^{-1}\log n - A(n\log n)^{-1} > 0, \qquad n > n_0(p).
\end{aligned}
$$

Thus since an harmonic function assumes its minimum value on the boundary, we have for $n > n_0(p)$

$$
(3.3)\quad \Re[zf_n'(z)/f_n(z)] > 0 \qquad \text{for} \qquad r \leqq 1 - (2p+2)n^{-1}\log n.
$$

To show that

$$
f_n(z) = z^p + a_{p+1}z^{p+1} + \cdots + a_nz^n
$$

is multivalent of order p in $|z| < 1 - (2p + 2)n^{-1} \log n$ we apply a theorem due to S. Ozaki (4) which states that if $f(z)$ is analytic in $|z| \leqq r$ and has p zeros there, none on the circumference, and if for some real α

$$\Re[e^{i\alpha} z f'(z)/f(z)] > 0 \qquad \text{on} \qquad |z| = r,$$

then $f(z)$ is multivalent of order p in the circle $|z| = r$. Applying the theorem to $f_n(z)$ we observe that $f_n(z)$ has no zeros on the boundary of the circle $|z| = 1 - (2p + 2)n^{-1} \log n$ since for $n > n_0(p)$

$$|f_n(z)| \geqq |f(z)| - |\rho_n(z)| > A.$$

Further $f_n(z)$ has exactly p zeros inside the given circle. This follows by the theorem of Rouché. On the circle

$$|f_n(z)| > A, \qquad |\rho_n(z)| < An^{-2}(\log n)^{-1}.$$

Therefore on the circle for $n > n_0(p)$ $|f_n(z)| > |\rho_n(z)|$, $f_n(z) \neq 0$. Hence by the theorem of Rouché $f_n(z)$ and $f_n(z) + \rho_n(z) = f(z)$ have the same number of zeros inside the circle. But since $f(z)$ has p zeros at the origin and is multivalent of order p in the unit circle it follows that $f(z)$ has exactly p zeros for $|z| < 1$. We have thus shown that the conditions of the theorem of S. Ozaki hold for $f_n(z)$ (with $\alpha = 0$). Thus $f_n(z)$ is multivalent of order p and star-like by (3.3) for $|z| < 1 - (2p + 2)n^{-1} \log n$. This completes the proof of Theorem A.

4. The partial sums of $z^p(1 - z)^{-2p}$

Let us consider the following particular multivalently star-like function of order p

$$f(z) = z^p(1 - z)^{-2p} = \sum_{k=p}^{\infty} \frac{(k + p - 1)!}{(k - p)!(2p - 1)!} z^k \tag{4.1}$$

and its n^{th} partial sum

$$f_n(z) = z^p + 2pz^{p+1} + \cdots + \frac{(n + p - 1)!}{(n - p)!(2p - 1)!} z^n. \tag{4.2}$$

We define R_n to be the radius of the largest circle $|z| = R_n$ within which $f_n(z)$ is multivalently star-like of order p. We shall show that for $n > n_0(p)$

$$R_n \geqq 1 - (2p + 1)n^{-1} \log n$$

$$\limsup_{n \to \infty} \left\{\frac{1 - R_n}{n^{-1} \log n}\right\} = 2p + 1. \tag{4.3}$$

We shall first establish by mathematical induction the following identity

$$f_n(z) = z^p(1 - z)^{-2p}\left[1 - \frac{(n + p)!}{(n - p)!(2p - 1)!} \int_0^z z^{n-p}(1 - z)^{2p-1} dz\right]. \tag{4.4}$$

The above formula is easily verified to be true for $n = p$. Let us now assume it correct for a value n and then show that this implies that the formula is also correct for the value $n + 1$. We first note an integration by parts gives

$$\frac{2p}{n-p+1}\int_0^z z^{n-p+1}(1-z)^{2p-1}\,dz$$
$$= -\frac{(1-z)^{2p}z^{n-p+1}}{n-p+1} + \int_0^z (1-z)(1-z)^{2p-1}z^{n-p}\,dz$$
$$= -\frac{(1-z)^{2p}z^{n-p+1}}{n-p+1} + \int_0^z z^{n-p}(1-z)^{2p-1}\,dz - \int_0^z z^{n-p+1}(1-z)^{2p-1}\,dz,$$
$$\frac{n+p+1}{n-p+1}\int_0^z z^{n-p+1}(1-z)^{2p-1}\,dz = \int_0^z z^{n-p}(1-z)^{2p-1}\,dz - \frac{z^{n-p+1}(1-z)^{2p}}{n-p+1}.$$

Using the last equality we obtain

$$f_{n+1}(z) = f_n(z) + \frac{(n+p)!}{(n-p+1)!(2p-1)!}z^{n+1}$$
$$= z^p(1-z)^{-2p}\left[1 - \frac{(n+p+1)!}{(n-p+1)!(2p-1)!}\right.$$
$$\left.\cdot\frac{n-p+1}{n+p+1}\left\{\int_0^z z^{n-p}(1-z)^{2p-1}\,dz - \frac{z^{n-p+1}(1-z)^{2p}}{n-p+1}\right\}\right]$$
$$= z^p(1-z)^{-2p}\left[1 - \frac{(n+p+1)!}{(n-p+1)!(2p-1)!}\int_0^z z^{n-p+1}(1-z)^{2p-1}\,dz\right].$$

Thus we have shown that the formula is correct for the value $n + 1$. This completes the proof of the formula.

Next we differentiate logarithmically the formula (4.4) and obtain

$$(4.5)\qquad zf_n'(z)/f_n(z) = p\left(\frac{1+z}{1-z}\right) - \frac{(n+p)!z^{n-p+1}(1-z)^{2p-1}}{\left\{(2p-1)!(n-p)! - (n+p)!\int_0^z z^{n-p}(1-z)^{2p-1}\,dz\right\}}.$$

For value of $|z| = r < 1$ for which

$$(2p-1)!(n-p)! - (n+p)!\int_0^r r^{n-p}(1+r)^{2p-1}\,dr > 0$$

we may write

$$(4.6)\qquad \Re[zf_n'(z)/f_n(z)] \geq p\left(\frac{1-r}{1+r}\right) - \frac{(n+p)!r^{n-p+1}(1+r)^{2p-1}}{(2p-1)!(n-p)! - (n+p)!\int_0^r r^{n-p}(1+r)^{2p-1}\,dr}.$$

Since $\int_0^z z^{n-p}(1-z)^{2p-1}\,dz$ is a polynomial with real coefficients whose signs alternate, it is seen that for $(n-p)$ odd equality is obtained in (4.6) when $z=-r$.

We shall then have

$$\Re[zf_n'(z)/f(z)] > 0$$

provided

$$\text{(i)}\quad (2p-1)!(n-p)!p(1-r)-(n+p)!p(1-r)\int_0^r r^{n-p}(1+r)^{2p-1}\,dr - (n+p)!r^{n-p+1}(1+r)^{2p}>0$$

and

$$\text{(ii)}\qquad (2p-1)!(n-p)!-(n+p)!\int_0^r r^{n+p}(1+r)^{2p-1}\,dr>0.$$

It is readily seen that the inequality (i) implies the inequality (ii).

Let the left-hand side of the inequality (i) be denoted by $T_n(r)$ and let $r = 1-\frac{\alpha}{n}$ where $\alpha=(2p+1)\log n$.

For $n>n_0(p)$

$$T_n(r) > (n-p)!(2p)!\tfrac{1}{2}(1-r) - (n+p)!p(1-r)r^{n-p+1}2^{2p-1}-(n+p)!2^{2p}r^{n-p+1}$$

$$\begin{aligned}\frac{T_n(r)}{(n-p)!} &> (2p+1)!(2n)^{-1}\log n \\ &\quad - (2n)^{2p}\cdot p\cdot 1\cdot\left(1-\frac{\alpha}{n}\right)^{1-p}e^{-\alpha}\cdot 2^{2p-1}-(2n)^{2p}\cdot 2^{2p}\left(1-\frac{\alpha}{n}\right)^{1-p}e^{-\alpha}\\ &> (2p+1)!(2n)^{-1}\log n - A(p)n^{2p}e^{-\alpha}\\ &\geqq (2p+1)!(2n)^{-1}\log n - A(p)n^{-1}\end{aligned}$$

where $A(p)$ is a constant depending upon p. Hence for $n>n_0(p)$

$$T_n\{1-(2p+1)n^{-1}\log n\}>0. \tag{4.7}$$

On the other hand if we denote by $A(p,\epsilon)$ a constant depending only upon p and ϵ, $0<\epsilon<1$, then for $n>n_0(p,\epsilon)$ when $r=1-(2p+1-\epsilon)n^{-1}\log n$ we have

$$\begin{aligned}\frac{T_n(r)}{(n-p)!} &< (2p)!(2p+1-\epsilon)(2n)^{-1}\log n - \frac{(n+p)!r^n}{(n-p)!}\\ &< (2p)!(2p+1-\epsilon)(2n)^{-1}\log n - n^{2p}\cdot A(p,\epsilon)n^{\epsilon-1-2p}\\ &\leqq (2p)!(2p+1-\epsilon)(2n)^{-1}\log n - A(p,\epsilon)n^{\epsilon-1}.\end{aligned}$$

Hence for $n > n_0(p, \epsilon)$

$$T_n\{1 - (2p + 1 - \epsilon)n^{-1} \log n\} < 0. \tag{4.8}$$

Thus we have shown that for $|z| \leqq 1 - (2p + 1)n^{-1} \log n$ and for $n > n_0(p)$

$$\Re[zf'_n(z)/f_n(z)] > 0 \tag{4.9}$$

and that the constant $(2p + 1)$ cannot be replaced by a smaller one for all $(n - p)$ odd.

A proof similar to that given at the end of the preceding section shows that the n^{th} partial sum for the power series of $z^p(1 - z)^{-2p}$ is multivalent of order p for $|z| < 1 - (2p + 1)n^{-1} \log n$. Thus we have for $n > n_0(p)$

$$R_n \geqq 1 - (2p + 1)n^{-1} \log n, \tag{4.10}$$

$$\limsup_{n \to \infty} \left\{\frac{1 - R_n}{n^{-1} \log n}\right\} = 2p + 1. \tag{4.11}$$

Rutgers University,
New Brunswick, N. J.

REFERENCES

1. M. S. Robertson, *On the theory of univalent functions*, Annals of Mathematics, vol. 37, 1936, pp. 374–408.
2. L. Bieberbach, *Lehrbuch der Funktionentheorie*, Berlin, vol. 2, 1931, pp. 71–83.
3. G. Pólya and G. Szegö, *Aufgaben und Lehrsatze aus der Analysis*, Berlin, vol. 1, 1925, p. 140.
4. S. Ozaki, *Some remarks on the univalency and multivalency of functions*, Science Reports of Tokyo Bunrika Daigaku, sect. A, 2, no. 32, 1934, pp. 41–55.

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

TWO-SIDED IDEALS AND CONGRUENCES IN THE RING OF BOUNDED OPERATORS IN HILBERT SPACE

J. W. CALKIN

(Received August 29, 1940)

Introduction

The developments of the present paper center around the observation that the ring $\mathfrak{B}$ of bounded everywhere defined operators in Hilbert space contains non-trivial two-sided ideals.[1] This fact, which has escaped all but oblique notice in the development of the theory of operators, is of course fundamental from the point of view of algebra and at the same time differentiates $\mathfrak{B}$ sharply from the ring of all linear operators over a unitary space with finite dimension number.

As examples of two-sided ideals in $\mathfrak{B}$ we may mention here the class of all operators A such that $\mathfrak{R}(A)$, the range of A, has a finite dimension number, the class of all operators of Hilbert-Schmidt type,[2] and the class $\mathfrak{T}$ of all totally continuous operators. Except for the ideal (0), every two-sided ideal in $\mathfrak{B}$ contains the first ideal mentioned, and except for the ideal $\mathfrak{B}$ itself, every two-sided ideal in $\mathfrak{B}$ is contained in the ideal $\mathfrak{T}$. Moreover, on the basis of the special spectral properties of the self-adjoint members of $\mathfrak{T}$, it is possible to characterize every two-sided ideal in $\mathfrak{B}$ very simply in terms of the spectra of its nonnegative self-adjoint elements; for both the formulation and the proof of this result, which together with the facts mentioned above is discussed in §1, the author is indebted to J. v. Neumann.

The restriction of our attention to those ideals in $\mathfrak{B}$ which are two-sided is basic for the points which we wish to develop; the two-sidedness compensates for the absence of commutativity in $\mathfrak{B}$ in such a way as to permit the construction of quotient rings by the standard methods of abstract algebra.[3] These rings, which are of course homomorphs of $\mathfrak{B}$ with respect to addition and multiplication, are also homomorphs of $\mathfrak{B}$ with respect to the operation *, and exhibit all of the formal properties of matrix algebras. This is established in §2, and there also various properties of the associated congruences in $\mathfrak{B}$ are discussed.

The remainder of the paper deals solely with the quotient ring $\mathfrak{B}/\mathfrak{T}$, where $\mathfrak{T}$ is the ideal of totally continuous operators, and the associated congruence in $\mathfrak{B}$. For essentially topological reasons, this is the only one of the quotient rings in

[1] An additive subset $\mathfrak{J}$ of $\mathfrak{B}$ is a left (right) ideal if it contains AB (BA) for all A in $\mathfrak{J}$ and B in $\mathfrak{B}$. If $\mathfrak{J}$ is both a left and right ideal, it is called two-sided. See references [1] and [19] at the end of the paper for the elementary properties of ideals.

[2] For a discussion in abstract terms of operators of this type (operators of "finite norm"), see reference [17], pp. 65–70.

[3] See, for example, [1], pp. 252–253, [19], pp. 56–57.

question which at present appears susceptible of deep analysis. For while there is in general no apparent way of introducing a topology in quotient rings over $\mathfrak{B}$, the ring $\mathfrak{B}/\mathfrak{T}$ is actually a complete metric space, the norm deriving very simply from the norm (i.e., bound) in $\mathfrak{B}$ itself (§3).

Moreover, it is even possible to interpret $\mathfrak{B}/\mathfrak{T}$ as an algebraic ring[4] of operators in a suitably defined complex Euclidean space $\mathfrak{L}$ whose dimension number is the cardinal number of the continuum.[5] Or, to put it differently, there exists in the ring of bounded everywhere-defined operators over $\mathfrak{L}$ a subset $\mathfrak{M}$ which is a $(+, \cdot, *)$-isomorphism of $\mathfrak{B}/\mathfrak{T}$. Furthermore, this correspondence is even an isometry, the norm of an element of $\mathfrak{B}/\mathfrak{T}$ being the bound of the corresponding element of $\mathfrak{M}$. These facts are all established in §§4, 5, and in §5, various properties of the algebraic ring $\mathfrak{M}$ are discussed.

Apart from its intrinsic interest, the analysis of the ring $\mathfrak{M}$ yields in a very simple way theorems of considerable depth concerning $\mathfrak{B}$ (§5). Of results of this sort, we shall here mention only a generalization of Weyl's classical theorem comparing the spectra of two self-adjoint operators with totally continuous difference [20].[6]

Before proceeding, we should like to point out various further developments which are suggested by the present paper and to which we propose to return at another time.

First, while the algebraic ring $\mathfrak{M}$ is not closed with respect to the weak topology for operators and so is not an operator ring in the sense of von Neumann [9, 11] one obtains such a ring by adjoining to $\mathfrak{M}$ its weak condensation points in the ring of bounded everywhere-defined operators in $\mathfrak{L}$. This "closure" $\mathbf{R}(\mathfrak{M})$ is of considerable interest from the point of view of the Murray-von Neumann theory of factors, especially since various preliminary results concerning $\mathbf{R}(\mathfrak{M})$ and $\mathfrak{M}'$ suggest that they may be factors of class III_∞.[7]

Second, while the factors of class II_1 and III_∞ of Murray and von Neumann are like those of class I_n, $n < \aleph_0$, simple rings, those of class II_∞ are not. The construction of quotient rings over such a factor is therefore possible and gives rise to various questions analogous to those concerning $\mathfrak{B}$ with which we deal here.[8]

Finally, we should like to point out that the maximal property of the ideal $\mathfrak{T}$ of totally continuous operators described above does not persist if one considers

[4] We use the term algebraic ring with reference to operators to denote a class closed with respect to the operations $+, \cdot, *$, and scalar multiplication. This is not a ring in the sense of von Neumann, [11], since no topological conditions are imposed.

[5] For the theory of complex Euclidean spaces of arbitrary dimension number, see references [6], [7], [13], [16].

[6] All numbers in brackets refer to the bibliography at the end of the paper.

[7] For the notion of factors, and their classification, see [9]; for a construction of factors of class III_∞, see [15].

[8] The class of members of a II_∞ factor which are "normed" in the sense of [15] is a two-sided ideal.

instead of Hilbert space, a space $\mathfrak{H}$ with dimension number greater than $\aleph_0$. For example, in such a space, the class of all operators A such that $\mathfrak{R}(A)$ contains no closed subspace with dimension number exceeding $\aleph_0$ is a non-trivial ideal. Moreover, if m is the dimension number of $\mathfrak{H}$, $m \geqq \aleph_0$, the class of all operators A which have the property that $\mathfrak{R}(A)$ contains no closed linear subspace of $\mathfrak{H}$ with dimension number m is a two-sided ideal different from $\mathfrak{B}$ and is identical with $\mathfrak{T}$ if $m = \aleph_0$. Furthermore, in every case, this ideal is maximal in the same sense in which $\mathfrak{T}$ is for $m = \aleph_0$, and as von Neumann has pointed out to the author, the spectral characterization of ideals can be extended in a satisfactory way to describe the general case. Finally, investigations which are still incomplete suggest that a considerable portion of the analysis of the present paper has a counterpart in every case.

1. Two-sided ideals in $\mathfrak{B}$

We proceed now to the characterization of all two-sided ideals in $\mathfrak{B}$.

THEOREM 1.1. *If $\mathfrak{I}$ is a left (right) ideal, the set $\mathfrak{I}^*$ of all adjoints A^* of elements A of $\mathfrak{I}$ is a right (left) ideal.*

Since $\mathfrak{I}^*$ is evidently closed with respect to addition, it is necessary only to show—when $\mathfrak{I}$ is a left ideal—that $\mathfrak{I}^*$ contains BA^* for all B in $\mathfrak{B}$ and A in $\mathfrak{I}$. But this follows at once from the relation $BA^* = (AB^*)^*$ and the fact that AB^* belongs to $\mathfrak{I}$ along with A. When $\mathfrak{I}$ is a right ideal, an analogous argument is valid.

THEOREM 1.2. *A necessary and sufficient condition that a left (right) ideal $\mathfrak{I}$ be two-sided is that $\mathfrak{I} = \mathfrak{I}^*$.*

The sufficiency of the condition is an immediate consequence of Theorem 1.1. Now suppose $\mathfrak{I}$ is two-sided, A an arbitrary member of $\mathfrak{I}$, $A = WB$ its canonical decomposition.[9] Then $A^* = BW^* = W^*AW^*$ is clearly in $\mathfrak{I}$ and $\mathfrak{I} = \mathfrak{I}^*$.

Since we now have no further direct concern with left or right ideals we shall refer to two-sided ideals merely as ideals.

THEOREM 1.3. *The class $\mathfrak{T}$ of all totally continuous operators in $\mathfrak{B}$ is an ideal.*

That $\mathfrak{T}$ is an additive class is obvious from the definition of a totally continuous operator; T is totally continuous if it takes every bounded set into a compact set. Moreover, by a well-known theorem, $\mathfrak{T} = \mathfrak{T}^*$.[10] Finally, since every member A of $\mathfrak{B}$ clearly takes bounded sets into bounded sets, $\mathfrak{T}$ is a left ideal. Hence by Theorem 2 it is an ideal.

Throughout the remainder of the paper $\mathfrak{T}$ has the same meaning as in Theorem 1.3.

THEOREM 1.4. *Let $\mathfrak{I}$ be an arbitrary ideal in $\mathfrak{B}$. Then either $\mathfrak{I} = \mathfrak{B}$ or $\mathfrak{I} \subseteq \mathfrak{T}$.*

The proof of this theorem is based on a characteristic property of totally continuous operators which the writer has noted elsewhere.[11] According to

[9] For the notion of canonical decomposition, see [9] and [12].
[10] [2], p. 100, Théorème 4.
[11] [3], Lemma 3.1.

this result a member T of $\mathfrak{B}$ is totally continuous if and only if every closed linear manifold in its range has a finite dimension number. Hence, if $\mathfrak{I}$ is an ideal which is not contained in $\mathfrak{T}$, $\mathfrak{I}$ contains an element A such that $\mathfrak{R}(A)$ contains a Hilbert space $\mathfrak{M}$. We denote by $\mathfrak{N}$ the manifold of zeros of A and by A_1 the transformation induced on $\mathfrak{H} \ominus \mathfrak{N}$ by A.[12] The transformation A_1 evidently possesses an inverse and the same range as A; moreover the fact that A_1 is bounded assures us that the set $\mathfrak{A} = A_1^{-1}\mathfrak{M}$ is closed. Thus $\mathfrak{A}$ is also a Hilbert space. Hence there exist in $\mathfrak{B}$ partially isometric operators[13] X, Y both with initial sets $\mathfrak{H}$, while $\mathfrak{R}(X) = \mathfrak{A}$, $\mathfrak{R}(Y) = \mathfrak{M}$. Therefore the operator $B = Y^*AX$ has domain and range identically $\mathfrak{H}$ and belongs to $\mathfrak{B}$. But $Bf = 0$ implies either $Xf = 0$, Xf in $\mathfrak{N}$, or AXf in $\mathfrak{H} \ominus \mathfrak{M}$, and all of these are impossible in view of the definition of X and Y. Hence B^{-1} exists and since it is closed with domain $\mathfrak{H}$, it belongs to $\mathfrak{B}$. Therefore $I = B^{-1}B$ belongs to $\mathfrak{I}$ and $\mathfrak{I} = \mathfrak{B}$.

THEOREM 1.5. *Let $\mathfrak{I}$ be an arbitrary ideal in $\mathfrak{B}$, $\mathfrak{I}_0$ the class of nonnegative definite self-adjoint transformations in $\mathfrak{I}$. Then $\mathfrak{I}_0$ is the class of all operators $B = (A^*A)^{\frac{1}{2}}$ such that A is an element of $\mathfrak{I}$, and if $\mathfrak{I}_1$ is an ideal containing $\mathfrak{I}_0$, then $\mathfrak{I}_1 \supseteqq \mathfrak{I}$.*

If B is in $\mathfrak{I}_0$, it is obvious that B is of the form described in the theorem. On the other hand, if A is an arbitrary element of $\mathfrak{I}$, $(A^*A)^{\frac{1}{2}} = B = W^*A$, where W is partially isometric; hence $(A^*A)^{\frac{1}{2}}$ belongs to $\mathfrak{I}$, and thus to $\mathfrak{I}_0$. Finally the relations $A = WB$, $B = (A^*A)^{\frac{1}{2}}$ assure us that $\mathfrak{I}$ is the smallest ideal containing $\mathfrak{I}_0$.

If $\mathfrak{I}$ is an ideal in $\mathfrak{B}$, we call the subset $\mathfrak{I}_0$ defined in Theorem 1.5 the positive part of $\mathfrak{I}$. In order to characterize those subsets of the positive part of $\mathfrak{B}$ which appear as the positive parts of ideals, we recall that every self-adjoint operator T in $\mathfrak{T}$ can be reduced to diagonal form; that is, for each such T there exists a complete orthonormal set $\{\varphi_n\}$ in $\mathfrak{H}$ and a sequence $\{\lambda_n\}$ such that $T\varphi_n = \lambda_n\varphi_n$, $n = 1, 2, \cdots$. Moreover, the sequence $\{\lambda_n\}$ is convergent to zero, and nonnegative if T is in the positive part of $\mathfrak{T}$. We call this sequence a characteristic sequence of T.

We next observe that if T belongs to the positive part of an ideal $\mathfrak{I}$ and $\{\psi_n\}$ is another complete orthonormal set in $\mathfrak{H}$, then the member A of $\mathfrak{T}$ defined by the equations $A\psi_n = \lambda_n\psi_n$, $n = 1, 2, \cdots$, also belongs to the positive part of $\mathfrak{I}$, since $A = UTU^{-1}$, where U is unitary. Hence, if $\mathfrak{I}$ is an arbitrary ideal in $\mathfrak{B}$, $\mathfrak{I} \subsetneqq \mathfrak{T}$ the set of all characteristic sequences $\{\lambda_n\}$ belonging to members of $\mathfrak{I}_0$ may without ambiguity be called the *spectral set* of $\mathfrak{I}$.

We now characterize intrinsically those subsets of the class of nonnegative sequences with limit zero which occur as spectral sets. As is indicated in the introduction, this result is due to J. v. Neumann.

DEFINITION 1.1. *Let $\mathfrak{X}$ denote the class of all infinite sequences of nonnegative*

[12] We regard A as a transformation between Hilbert spaces in the sense of [8].

[13] For this concept, see [9], Definition 4.3.1.

numbers which converge to zero. A subset $\mathfrak{J}$ *of* $\mathfrak{T}$ *is called an ideal set if it has the following properties*:

(i) *If* $\{\lambda_n\}$ *is in* $\mathfrak{J}$ *and* π *denotes an arbitrary permutation of the positive integers,* $\{\lambda_{\pi(n)}\}$ *is in* $\mathfrak{J}$;

(ii) *if* $\{\lambda_n\}$ *and* $\{\mu_n\}$ *are in* $\mathfrak{J}$, *so is* $\{\lambda_n + \mu_n\}$;

(iii) *if* $\{\lambda_n\}$ *is in* $\mathfrak{J}$, $\{\mu_n\}$ *in* $\mathfrak{T}$ *and the inequality* $\lambda_n \geqq \mu_n$ *holds for all n, then* $\{\mu_n\}$ *is in* $\mathfrak{J}$.

Our object now is to prove that every spectral set is an ideal set, and conversely. We require first two lemmas concerning ideal sets.

LEMMA 1.1. *Let* $\mathfrak{J}$ *be an ideal set,* $\{\lambda_n\}$ *an element of* $\mathfrak{J}$ *with infinitely many terms different from zero. Then the subsequence of positive terms of* $\{\lambda_n\}$ *belongs to* $\mathfrak{J}$.

We distinguish two cases according as $\{\lambda_n\}$ contains a finite or an infinite number of zeros. In view of condition (i) of Definition 1, we can assume in the first case $\lambda_1 = \lambda_2 = \cdots = \lambda_N = 0$ and $0 < \lambda_{n+1} \leqq \lambda_n$, $n = N + 1, N + 2, \cdots$. We then have to show that $\{\lambda_{N+n}\}$ belongs to $\mathfrak{J}$. Let $\{\mu_n\}$ be the sequence defined by the equations

$$\mu_n = \lambda_{N+n}, \qquad n = 1, 2, \cdots, N,$$
$$\mu_n = 0, \qquad n = N + 1, N + 2, \cdots.$$

Then $\{\mu_n\}$ is dominated by a permutation of $\{\lambda_n\}$ and hence belongs to $\mathfrak{J}$, by conditions (i) and (iii) of Definition 1. Moreover, by condition (ii), $\{\lambda_n + \mu_n\}$ belongs to $\mathfrak{J}$. But we have $\lambda_{N+n} \leqq \lambda_n + \mu_n$, $n = 1, 2, \cdots$, and hence $\{\lambda_{N+n}\}$ belongs to $\mathfrak{J}$.

Now suppose $\{\lambda_n\}$ contains an infinite number of zeros. Again invoking (i), we assume

$$\lambda_{2k} \geqq \lambda_{2(k+1)} > 0, \qquad \lambda_{2k-1} = 0, \qquad k = 1, 2, \cdots.$$

We must then show that $\{\lambda_{2n}\}$ belongs to $\mathfrak{J}$. We define a sequence $\{\mu_n\}$ by the equations

$$\mu_{2k} = \lambda_{2k-1}, \qquad \mu_{2k-1} = \lambda_{2k}, \qquad k = 1, 2, \cdots.$$

Then $\{\mu_n\}$ is a permutation of $\{\lambda_n\}$ and hence belongs to $\mathfrak{J}$. But then $\{\lambda_n + \mu_n\}$ belongs to $\mathfrak{J}$, and we have

$$\lambda_{2n} \leqq \lambda_n + \mu_n$$

Thus, by (iii), $\{\lambda_{2n}\}$ belongs to $\mathfrak{J}$ as we wished to prove.

LEMMA 1.2. *Let* $\mathfrak{J}$ *be an ideal set,* $\{\lambda_n\}$ *an element of* $\mathfrak{J}$ *consisting solely of positive terms. Then any sequence* $\{\mu_n\}$ *which contains* $\{\lambda_n\}$ *as a subsequence and which, except for this subsequence, consists solely of zeros, is a member of* $\mathfrak{J}$.

We define

$$\lambda_{2k-1}^{(1)} = 0, \qquad \lambda_{2k}^{(1)} = \lambda_{2k}, \qquad k = 1, 2, \cdots,$$
$$\lambda_{2k-1}^{(2)} = \lambda_{2k-1}, \qquad \lambda_{2k}^{(2)} = 0, \qquad k = 1, 2, \cdots.$$

Then $\{\lambda_n^{(1)}\}$ and $\{\lambda_n^{(2)}\}$ are both dominated by $\{\lambda_n\}$ and therefore belong to $\mathfrak{J}$. Letting j_k denote the integer immediately preceding $k/3$, we now define two permutations π_1, π_2 by the equations,

$$\pi_1(k + j_k) = 2k - 1, \qquad \pi_1(4k) \qquad = 2k, \qquad k = 1, 2, \cdots,$$

$$\pi_2(4k - 2) = 2k - 1, \qquad \pi_2(k + j_{k+2}) = 2k, \qquad k = 1, 2, \cdots.$$

The effect of the first of these permutations is to take the set of all integers divisible by 4 into the set of all even integers, preserving order; and to take the set of all integers not divisible by 4 into the set of all odd integers, again preserving order. The effect of the second is to take all even integers not divisible by 4 into the set of all odd integers, and the complementary set of integers into the set of all even integers, order being preserved in both cases. Hence the sequence

$$\{\nu_n\} = \{\lambda_{\pi_1(n)}^{(1)} + \lambda_{\pi_2(n)}^{(2)}\}$$

which clearly belongs to $\mathfrak{J}$, is related to the sequence $\{\lambda_n\}$ by the equations

$$\nu_{2k} = \lambda_k, \qquad \nu_{2k-1} = 0, \qquad k = 1, 2, \cdots.$$

But any sequence $\{\mu_n\}$ derived from $\{\lambda_n\}$ in the manner described in the theorem, and containing an infinite number of zero terms, is a permutation of $\{\nu_n\}$ and therefore belongs to $\mathfrak{J}$.

Thus, to complete the proof, we have only to dispose of the case that $\{\mu_n\}$ contains only a finite number of zero terms. In this case it is convenient to assume that $\{\lambda_n\}$ is monotone. We can then complete the proof by showing that the sequence $\{\mu_n\}$ defined by the equations

$$\mu_n = 0, \qquad n = 1, 2, \cdots, N - 1, \qquad \mu_{N+n-1} = \lambda_n, \qquad n = 1, 2, \cdots,$$

belongs to $\mathfrak{J}$. To do this we set

$$\nu_n = \lambda_1, \qquad n = 1, 2, \cdots, N, \qquad \nu_n = \lambda_2, \qquad n = N + 1, \\ N + 2, \cdots, 2N, \cdots.$$

Then, from the validity of the lemma in the case of an infinite number of zeros and property (ii) of ideal sets we can conclude that $\{\nu_n\}$ is in $\mathfrak{J}$. But, since we clearly have

$$\mu_n \leqq \nu_n$$

it follows that $\{\mu_n\}$ is in $\mathfrak{J}$.

THEOREM 1.6. *Let $\mathcal{I}$ be an ideal in $\mathfrak{B}$, $\mathcal{I} \subsetneqq \mathcal{T}$, and let $\mathfrak{J}$ be its spectral set. Then $\mathfrak{J}$ is an ideal set. Conversely, if $\mathfrak{J}$ is an arbitrary ideal set in $\mathfrak{T}$, there exists an ideal in $\mathfrak{B}$ whose spectral set is $\mathfrak{J}$. This correspondence between ideals in $\mathfrak{B}$ and ideal sets in $\mathfrak{T}$ is an isomorphism with respect to the relation $\subset$.*

If $\mathcal{I}$ is a two-sided ideal in $\mathfrak{B}$, $\mathcal{I} \subsetneqq \mathcal{T}$, we consider an arbitrary member A of $\mathcal{I}_0$ in diagonal form; $A\varphi_n = \lambda_n\varphi_n$, $n = 1, 2, \cdots$, where $\{\varphi_n\}$ is a complete orthonormal set in $\mathfrak{H}$. Then $\{\lambda_n\}$ is in the spectral set $\mathfrak{J}$ of $\mathcal{I}$, and since A can also

be put in diagonal form with reference to any permutation of the sequence $\{\varphi_n\}$, all permutations of $\{\lambda_n\}$ clearly belong to $\mathfrak{J}$ too.

Now let B be any other member of $\mathcal{J}_0$, also in diagonal form; $B\psi_n = \mu_n\psi_n$, $n = 1, 2, \cdots$, and let U be the unitary operator defined on $\{\varphi_n\}$ by the equations $U\varphi_n = \psi_n$, $n = 1, 2, \cdots$. Then $A + U^{-1}BU$ is in $\mathcal{J}_0$ and

$$(A + U^{-1}BU)\varphi_n = (\lambda_n + \mu_n)\varphi_n, \qquad n = 1, 2, \cdots.$$

Thus $\{\lambda_n + \mu_n\}$ belongs to $\mathfrak{J}$.

Finally, let $\{\mu_n\}$ be a sequence such that $\mu_n \leqq \lambda_n$, $n = 1, 2, \cdots$. We define an operator B_0 on $\{\varphi_n\}$ by the equations

$$B_0\varphi_n = (\mu_n/\lambda_n)\varphi_n, \qquad \lambda_n \neq 0,$$
$$B_0\varphi_n = 0 \qquad\qquad , \qquad \lambda_n = 0.$$

Then B_0 has a closed linear extension B which belongs to $\mathfrak{B}$. Thus AB belongs to $\mathcal{J}$. But $AB\varphi_n = \mu_n\varphi_n$, $n = 1, 2, \cdots$, and therefore AB belongs to $\mathcal{J}_0$, $\{\mu_n\}$ to $\mathfrak{J}$.

We turn now to the converse part of the theorem, denoting by $\mathfrak{J}$ an arbitrary ideal set in $\mathfrak{T}$. We designate by $\mathcal{J}$ the set of all totally continuous operators A such that a characteristic sequence of $(A^*A)^{\frac{1}{2}}$ belongs to $\mathfrak{J}$. Since, if $\mathcal{J}$ is an ideal, $\mathfrak{J}$ is obviously its spectral set, we have only to show that $\mathcal{J}$ is an ideal.

We begin by considering an arbitrary pair of operators A and B of $\mathcal{J}$, with the object of showing that $A + B$ belongs to $\mathcal{J}$. To this end we consider the operators

$$D_1 = (A^*A)^{\frac{1}{2}}, \qquad D_2 = (B^*B)^{\frac{1}{2}}, \qquad D = (A^* + B^*)(A + B)^{\frac{1}{2}}$$

and characteristic sequences $\{\lambda_n\}$, $\{\mu_n\}$, $\{\nu_n\}$ of D_1, D_2, D, respectively. We have then to show that $\{\nu_n\}$ belongs to $\mathfrak{J}$.

For convenience, we assume that the positive terms of each sequence are arranged in montone order; in addition, if any one of the sequences contains only a finite number of positive terms, we assume that the sequence itself is monotone. This assumption is clearly not restrictive, in view of condition (i) of Definition 1. Now let $\{\lambda_n'\}$ be identical with the subsequence of positive terms of $\{\lambda_n\}$ if that subsequence is infinite, identical with $\{\lambda_n\}$ itself in the alternative case, and let $\{\mu_n'\}$ and $\{\nu_n'\}$ be defined in the same way with reference to $\{\mu_n\}$ and $\{\nu_n\}$ respectively. Then, by Lemma 1.1, $\{\lambda_n'\}$ and $\{\mu_n'\}$ belong to $\mathfrak{J}$; and by Lemma 1.2, $\{\nu_n\}$ belongs to $\mathfrak{J}$ if $\{\nu_n'\}$ does. Thus we can prove that $A + B$ belongs to $\mathcal{J}$ by showing that $\{\nu_n'\}$ belongs to $\mathfrak{J}$.

To establish the latter result, we note first the relation

$$2(A^*A + B^*B) - (A^* + B^*)(A + B) = (A^* - B^*)(A - B),$$

which implies that $2(D_1^2 + D_2^2) - D^2$ is nonnegative definite. Holding this fact in reserve, we then recall the theorem of Courant [4],[14] characterizing the

[14] The paper cited deals only with integral operators with continuous kernels. However, the more general result required here is readily obtained.

sequence $\{\nu_n'^2\}$ associated with the operator D^2 in terms of maxima and minima. For our purposes, we may state this result as follows:

$$\nu_n'^2 = \underset{\dim(\mathfrak{M}) \leqq n-1}{\mathrm{Min}} \ \underset{\substack{f \in \mathfrak{H} \ominus \mathfrak{M} \\ |f|=1}}{\mathrm{Max}} (D^2 f, f) \tag{1}$$

where dim $(\mathfrak{M})$ is the dimension number of $\mathfrak{M}$. Similarly,

$$\lambda_n'^2 = \underset{\dim(\mathfrak{N}) \leqq n-1}{\mathrm{Min}} \ \underset{\substack{f \in \mathfrak{H} \ominus \mathfrak{N} \\ |f|=1}}{\mathrm{Max}} (D_1^2 f, f), \tag{2}$$

$$\mu_n'^2 = \underset{\dim(\mathfrak{P}) \leqq n-1}{\mathrm{Min}} \ \underset{\substack{f \in \mathfrak{H} \ominus \mathfrak{P} \\ |f|=1}}{\mathrm{Max}} (D_2^2 f, f). \tag{3}$$

On the basis of these relations we are able to conclude that the inequality

$$\nu_n'^2 \leqq 2(\lambda_j'^2 + \mu_k'^2) \tag{4}$$

is valid provided we have $j + k \leqq n + 1$. For if $\nu_n'^2 > 2(\lambda_j'^2 + \mu_k'^2)$ holds, we have, in view of (2) and (3),

$$\nu_n'^2/2 > \underset{\dim(\mathfrak{N}) \leqq j-1}{\mathrm{Min}} \ \underset{f \in \mathfrak{H} \ominus \mathfrak{N}}{\mathrm{Max}} (D_1^2 f, f) + \underset{\dim(\mathfrak{P}) \leqq k-1}{\mathrm{Min}} \ \underset{f \in \mathfrak{H} \ominus \mathfrak{P}}{\mathrm{Max}} (D_2^2 f, f)$$

and this implies

$$\nu_n'^2/2 > \underset{\dim(\mathfrak{N}+\mathfrak{P}) \leqq j+k-2}{\mathrm{Min}} \left[\underset{f \in \mathfrak{H} \ominus \mathfrak{N}}{\mathrm{Max}} (D_1^2 f, f) + \underset{f \in \mathfrak{H} \ominus \mathfrak{P}}{\mathrm{Max}} (D_2^2 f, f) \right]$$

which in turn implies

$$\nu_n'^2/2 > \underset{\dim(\mathfrak{N}+\mathfrak{P}) \leqq j+k-2}{\mathrm{Min}} \left[\underset{f \in \mathfrak{H} \ominus (\mathfrak{N} \oplus \mathfrak{P})}{\mathrm{Max}} (D_1^2 f + D_2^2 f, f) \right],$$

f being restricted in every case to satisfy $|f| = 1$. But then, since $D_1^2 + D_2^2 - D^2/2$ is nonnegative definite, we have

$$\nu_n'^2 > \underset{\dim(\mathfrak{M}) \leqq j+k-2}{\mathrm{Min}} \ \underset{\substack{f \in \mathfrak{H} \ominus \mathfrak{M} \\ |f|=1}}{\mathrm{Max}} (D^2 f, f)$$

which contradicts (1) unless $j + k - 2$ is greater than $n - 1$, or unless $j + k$ exceeds $n + 1$. Thus (4) holds for $j + k \leqq n + 1$, as stated, and we have

$$\nu_n' \leqq 2(\lambda_j' + \mu_k') \qquad \text{for } j + k \leqq n + 1. \tag{5}$$

Now let r_n be $n/2$ if n is even, $(n + 1)/2$ if n is odd. Then, from (5) we have

$$\nu_n' \leqq 2(\lambda_{r_n}' + \mu_{r_n}') \tag{6}$$

Moreover, the sequences $\{\lambda_n''\}$, $\{\mu_n''\}$ defined by the equations

$$\lambda_{2k-1}'' = 0, \qquad \lambda_{2k}'' = \lambda_k', \qquad k = 1, 2, \cdots,$$

$$\mu_{2k-1}'' = 0, \qquad \mu_{2k}'' = \mu_k', \qquad k = 1, 2, \cdots,$$

belong to $\mathfrak{J}$ by Lemma 1.2; and since

$$\lambda_{r_n}' = \lambda_n'' + \lambda_{\pi(n)}'', \qquad \mu_{r_n} = \mu_n'' + \pi_{\pi(n)}'', \qquad n = 1, 2, \cdots$$

where $\pi(2k-1) = 2k$, $\pi(2k) = 2k-1$, it follows from conditions (i) and (iii) of Definition 1.1 that the sequences $\{\lambda'_{r_n}\}$ and $\{\mu'_{r_n}\}$ both belong to $\mathfrak{J}$. But then, by (6) and conditions (ii) and (iii) of our definition, $\{\nu'_n\}$ belongs to $\mathfrak{J}$, and this completes the proof of our assertion that $A + B$ belongs to $\mathcal{I}$.

Now let A be an arbitrary element of $\mathcal{I}$, $D_1 = (A^*A)^{\frac{1}{2}}$, X an arbitrary element of $\mathcal{B}$, $D_2^2 = (A^*X^*XA)^{\frac{1}{2}}$. Since A is in $\mathcal{I}$, XA and D_2 are also. Let $\{\lambda_n\}$, $\{\mu_n\}$, $\{\lambda'_n\}$, $\{\mu'_n\}$ have the same meanings as above with reference to D_1 and D_2. Then $\{\lambda'_n\}$ is in $\mathfrak{J}$ and we can show that XA is in $\mathcal{I}$ by showing that $\{\mu'_n\}$ is in $\mathfrak{J}$; the arguments here are the same as above.

To establish the latter, we again apply the theorem of Courant;

$$\lambda_n'^2 = \min_{\dim(\mathfrak{M}) \leq n-1} \max_{\substack{f \in \mathfrak{H} \ominus \mathfrak{M} \\ |f|=1}} (D_1^2 f, f),$$

$$\mu_n'^2 = \min_{\dim(\mathfrak{M}) \leq n-1} \max_{\substack{f \in \mathfrak{H} \ominus \mathfrak{M} \\ |f|=1}} (D_2^2 f, f).$$

But $(D_1^2 f, f) = (Af, Af)$, $(D_2^2 f, f) = (XAf, XAf)$ and (XAf, XAf) is bounded by $N^2(Af, Af)$ for some integer N. Thus we have

$$\mu'_n \leqq N\lambda'_n,$$

and as $\{N\lambda'_n\}$ is clearly in $\mathfrak{J}$, so also is $\{\mu'_n\}$. Hence XA belongs to $\mathcal{I}$.

Thus $\mathcal{I}$ is a left ideal and in view of Theorem 1.2, we can show that $\mathcal{I}$ is two-sided by showing that $\mathcal{I}$ is closed with respect to the operation *. But this is an immediate consequence of Lemmas 1.1 and 1.2 and the well-known fact that $(A^*A)^{\frac{1}{2}}$ and $(AA^*)^{\frac{1}{2}}$ have the same positive characteristic values, each with the same multiplicity.[15]

The concluding assertion of the theorem is obvious.

THEOREM 1.7. *Let $\mathcal{F}$ denote the class of all operators A in $\mathcal{B}$ such that $\mathfrak{R}(A)$ has a finite dimension number. Then $\mathcal{F}$ is a two-sided ideal in $\mathcal{B}$. If $\mathcal{I}$ is an arbitrary two-sided ideal, in $\mathcal{B}$, then $\mathcal{I} = (0)$ or $\mathcal{I} \supseteqq \mathcal{F}$.*

To establish the first assertion, we have only to note that the class $\mathfrak{F}$ of all sequences in $\mathfrak{T}$ with only a finite number of terms different from zero constitutes an ideal set. To establish the second we first observe that any ideal set different from the one containing only the sequence all of whose terms are zero, contains a sequence $\{\lambda_n\}$ with $\lambda_1 \neq 0$, $\lambda_n = 0$ for $n \neq 1$, by virtue of (i) and(iii). Hence, by (ii) and (iii) it contains all sequences of this sort and hence, by (i) and (ii), contains $\mathfrak{F}$.

It is worthwhile to observe here that the effect of Theorem 1.6 is to establish an isomorphism between the lattice L_1 of two-sided ideals in $\mathcal{B}$ and a certain sub-lattice L_2 of the lattice of ideals in the ring $\mathfrak{B}$ of all bounded sequences of complex numbers. To make this clear we require two preliminary results concerning $\mathfrak{B}$; the first theorem is the analogue for $\mathfrak{B}$ of Theorem 1.5 for $\mathcal{B}$ the second is analogous to Theorems 1.3 and 1.4.

[15] By [12], Satz 7, for example.

THEOREM 1.8. *Let $\mathfrak{B}$ be the ring of all bounded sequences of complex numbers, $\mathfrak{J}$ an ideal in $\mathfrak{B}$. Let $\mathfrak{J}_0$ be the set of all sequences $\{|\lambda_n|\}$ such that $\{\lambda_n\}$ is in $\mathfrak{J}$. Then $\mathfrak{J}$ contains $\mathfrak{J}_0$, and if $\mathfrak{J}_1$ is an ideal in $\mathfrak{B}$ containing $\mathfrak{J}_0$, then $\mathfrak{J}_1 \supseteqq \mathfrak{J}$.*

The proof is straightforward and is left to the reader.

THEOREM 1.9. *Let L_2 be the lattice of all ideals $\mathfrak{J}$ in $\mathfrak{B}$ which satisfy the following condition: if $\{a_n\}$ is in $\mathfrak{J}$, and π is a permutation of the positive integers $\{a_{\pi(n)}\}$ is in $\mathfrak{J}$. Then the set $\mathfrak{T}$ of sequences convergent to zero belongs to L_2, and if $\mathfrak{J}$ is an arbitrary element of L_2, either $\mathfrak{J} = \mathfrak{B}$ or $\mathfrak{J} \subseteq \mathfrak{T}$.*

The first assertion is obvious. Now if $\mathfrak{J}$ is a member of L_2 which is not contained in $\mathfrak{T}$, $\mathfrak{J}$ contains a sequence $\{a_n\}$ which has a subsequence $\{a_{n_k}\}$ such that $\{a_{n_k}^{-1}\}$ is bounded, and which converges to a number different from zero. Furthermore, the sequence $\{b_n\}$ with $b_{n_k} = a_{n_k}$, $b_n = 0$, $n \neq n_k$ can be written $\{b_n\} = \{c_n\}\{a_n\}$, where $c_{n_k} = 1$, $c_n = 0$, $n \neq n_k$, and hence $\{b_n\}$ belongs to $\mathfrak{J}$. We now write the sequence $\{b_n\}$ in the form $\{b_n\} = \{\rho_n e^{i\theta_n}\}$, $\rho_n > 0$, and observe that $\{\rho_n\}$ also belongs to $\mathfrak{J}$ by Theorem 1.8. But then by the same sort of argument that was used to prove Lemma 1.1, we can show that $\{\rho_{n_k}\}$ belongs to $\mathfrak{J}$, and hence $\{a_{n_k}\}$ does also. Hence $\{e_k\} = \{a_{n_k}\}\{a_{n_k}^{-1}\}$ belongs to $\mathfrak{J}$ and $e_k = 1$, $k = 1, 2, \cdots$. Thus $\mathfrak{J} = \mathfrak{B}$.

THEOREM 1.10. *Let $\mathfrak{J}$ be a member of L_2, $\mathfrak{J}_0$ the subset of $\mathfrak{J}$ defined in Theorem 1.8. Then either $\mathfrak{J}_0 = \mathfrak{B}_0$, where $\mathfrak{B}_0$ the set of all sequences of nonnegative numbers in $\mathfrak{B}$, or every sequence in $\mathfrak{J}_0$ is convergent to zero and $\mathfrak{J}_0$ is an ideal set in the sense of Definition* 1.1.

THEOREM 1.10 follows at once from Theorems 1.8 and 1.9.

THEOREM 1.11. *Let L_1 be the lattice of all two-sided ideals in $\mathcal{B}$, let the set $\mathfrak{B}_0$ be called the spectral set of $\mathcal{B}$, and let the class of ideal sets be extended to include $\mathfrak{B}_0$. Then L_1 is lattice-isomorphic to the extended class of ideal sets, each member of L_1 corresponding under this isomorphism to its spectral set. Similarly, the lattice L_2 is lattice-isomorphic to the extended class of all ideal sets, each member $\mathfrak{J}$ of L_2 corresponding to its $\mathfrak{J}_0$. Thus L_1 and L_2 are lattice-isomorphic and under this isomorphism $\mathcal{B}$ corresponds to $\mathfrak{B}$, $\mathcal{T}$ to the class $\mathfrak{T}$ of all sequences convergent to zero, $\mathcal{F}$ to the class $\mathfrak{F}$ of all sequences with only a finite number of terms different from zero.*

Theorem 1.11 is obvious on the basis of preceding results and we omit the proof. Before proceeding, however, we wish to make the following observations: The ring $\mathfrak{B}$ can be imbedded in $\mathcal{B}$ in a very simple way; we have merely to choose a complete orthonormal set $\{\varphi_n\}$ in $\mathfrak{H}$ and identify the element $\{a_n\}$ of $\mathfrak{B}$ with the closed linear operator A in $\mathfrak{H}$ which is defined on $\{\varphi_n\}$ by the equations $A\varphi_n = a_n\varphi_n$, $n = 1, 2, \cdots$. Moreover, if we consider $\mathfrak{B}$ in terms of this identification, each two-sided ideal $\mathcal{J}$ in $\mathcal{B}$ corresponds under the isomorphism of Theorem 1.11 merely to its intersection with $\mathfrak{B}$. This suggests an alternative attack on the problem solved by Theorem 1.6; however, in so far as we can determine, it is not possible to devise any essentially different proof of that theorem on this basis.

2. Congruences in $\mathfrak{B}$

We pass now to the study of congruences modulo an ideal $\mathfrak{I}$ in $\mathfrak{B}$. Following the standard procedure of abstract algebra, we consider the class $\mathfrak{B}/\mathfrak{I}$ whose elements α, β, $\cdots$ are the residue classes of $\mathfrak{B}$ with respect to $\mathfrak{I}$; by definition, two members A and B of $\mathfrak{B}$ belong to the same element α of $\mathfrak{B}/\mathfrak{I}$ if and only if $A - B$ is in $\mathfrak{I}$. If α and β are arbitrary elements of $\mathfrak{B}/\mathfrak{I}$, we define $\alpha + \beta$ as the class of all elements $A + B$ of $\mathfrak{B}$ such that A is in α, B in β; similarly, we define $\alpha\beta$ as the class of all AB in $\mathfrak{B}$ such that A is in α, B in β. Then, from the general theorem[16] which is controlling in such situations, we have

THEOREM 2.1. *If α and β are elements of $\mathfrak{B}/\mathfrak{I}$, so also are $\alpha + \beta$ and $\alpha\beta$. With addition and multiplication defined in this way $\mathfrak{B}/\mathfrak{I}$ is a ring; that is to say, $\mathfrak{B}/\mathfrak{I}$ is a commutative group with respect to the operation $+$, and further, the following formal laws are satisfied:*

$$(\alpha\beta)\gamma = \alpha(\beta\gamma), \qquad \alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma, \qquad (\beta + \gamma)\alpha = \beta\alpha + \gamma\alpha.$$

Moreover, $\mathfrak{B}/\mathfrak{I}$ possesses a unit.

It may be noted that except in the case $\mathfrak{I} = \mathfrak{B}$, the subclass of $\mathfrak{B}/\mathfrak{I}$ each of whose elements contains a scalar multiple of the identity in $\mathfrak{B}$, is isomorphic to the class of scalar multiples of the identify in $\mathfrak{B}$, since no two of these elements of $\mathfrak{B}$ can have difference in $\mathfrak{I}$ unless they are identical. It is convenient therefore to use italic letters for these elements as well as for the corresponding elements of $\mathfrak{B}$. In addition we shall use the symbol 1 for the unit in $\mathfrak{B}/\mathfrak{I}$; that is, for the element of $\mathfrak{B}/\mathfrak{I}$ whose members have the form $I + T$, where I is the identity in $\mathfrak{B}$ and T belongs to $\mathfrak{I}$.

It is worth pointing out here that the ring $\mathfrak{B}/\mathfrak{I}$, $\mathfrak{I} \neq \mathfrak{B}$, is certainly non-commutative; to verify this one needs only to consider two orthogonal projections E and F whose ranges are Hilbert spaces and whose sum is the identity, a partially isometric operator W which maps E on F, and the operator $W^*W - WW^* = E - F$ which belongs to no ideal except $\mathfrak{B}$ itself. Later we shall show that the center of $\mathfrak{B}/\mathfrak{I}$, $\mathfrak{I} \neq \mathfrak{B}$, is the set of all scalar multiples of unity (Theorem 2.9).

THEOREM 2.2. *If α is an arbitrary element of $\mathfrak{B}/\mathfrak{I}$, the class α^* of all members A^* of $\mathfrak{B}$ such that A is in α is in $\mathfrak{B}/\mathfrak{I}$ also. The operation $*$ so defined in $\mathfrak{B}/\mathfrak{I}$ obeys the following laws:*

$$\alpha^{**} = \alpha, \qquad (\alpha + \beta)^* = \alpha^* + \beta^*, \qquad (\alpha\beta)^* = \beta^*\alpha^*.$$

That α^* is in $\mathfrak{B}/\mathfrak{I}$ follows at once from the fact that $\mathfrak{I}^* = \mathfrak{I}$; and the three laws stated in the theorem are readily verified on the basis of their validity in $\mathfrak{B}$.

Thus we see that the rings $\mathfrak{B}/\mathfrak{I}$ have all of the formal properties of matrix algebras and are homomorphs of $\mathfrak{B}$ with respect to the operations $+$, $\cdot$, $*$, $a\cdot$;

[16] [1], pp. 252–253. The missing details necessary for our purposes are readily supplied. Cf. the discussion of the commutative case in [19].

moreover, they are of course the only homomorphs of $\mathfrak{B}$ with respect to these operations.

It is now desirable to consider these homomorphisms with respect to the following important notions in operator theory; for an operator to be self-adjoint, to be idempotent, to be partially isometric, to be unitary. Hence we are led to define these concepts in $\mathfrak{B}/\mathfrak{I}$ without explicit reference to their meanings in $\mathfrak{B}$.

DEFINITION 2.1. *An element α of $\mathfrak{B}/\mathfrak{I}$ is called self-adjoint if $\alpha = \alpha^*$; a self-adjoint element ϵ of $\mathfrak{B}/\mathfrak{I}$ is called idempotent if $\epsilon^2 = \epsilon$; an element ω of $\mathfrak{B}/\mathfrak{I}$ is called partially isometric if $\omega^*\omega = \epsilon$ is idempotent, unitary if $\omega^*\omega = \omega\omega^* = 1$.*

It is easy to see that under the homomorphisms $\mathfrak{B} \to \mathfrak{B}/\mathfrak{I}$ the image of every-self adjoint operator is self-adjoint and that analogous assertions hold for projections, partially isometric operators, and unitary operators. We shall now show that with reference to the first two of these concepts the converse statements are also true.

THEOREM 2.3. *If α is a self-adjoint element of $\mathfrak{B}/\mathfrak{I}$, α contains a self-adjoint member of $\mathfrak{B}$, and conversely.*

Let α be self-adjoint, A an element of α. Then $A^* - A$ is in $\mathfrak{I}$ and hence $A + (A^* - A)/2 = (A + A^*)/2$ is in α. The converse, as we have already noted, is obvious.

THEOREM 2.4. *If ϵ is an idempotent element of $\mathfrak{B}/\mathfrak{I}$, there exists a projection E in $\mathfrak{B}$ which belongs to ϵ, and conversely.*

The theorem is obvious for $\mathfrak{I} = \mathfrak{B}$; we assume therefore $\mathfrak{I} \subseteq \mathfrak{T}$. By Theorem 2.3, ϵ contains a self-adjoint transformation A, and since ϵ is idempotent, $A^2 - A = A(A - I)$ is in $\mathfrak{I}$ and thus in $\mathfrak{T}$. Hence $A^2 - A$ can be reduced to diagonal form, and therefore A can also; $A\varphi_n = \lambda_n\varphi_n$, $n = 1, 2, \cdots$, where $\{\varphi_n\}$ is a complete orthonormal set in $\mathfrak{H}$. But then it follows that $\{\lambda_n\}$ contains a subsequence $\{\lambda_n^{(0)}\}$ convergent to zero, and such that the remaining terms of $\{\lambda_n\}$ form a subsequence, say $\{\lambda_n^{(1)}\}$, convergent to 1, since under any other circumstances $A(A - I)$ would fail to be totally continuous. Moreover, we can clearly assume that $\{\lambda_n^{(1)}\}$ contains no terms with the value zero and that $\{\lambda_n^{(0)}\}$ contains no terms with the value unity.

Now let $\mathfrak{M}_0$ be the subspace of $\mathfrak{H}$ determined by the characteristic elements of A corresponding to terms of $\{\lambda_n^{(0)}\}$, $\mathfrak{M}_1$ the subspace determined by the other characteristic elements of A, E_0 and E_1 the projections with ranges $\mathfrak{M}_0$ and $\mathfrak{M}_1$, respectively. We shall show that E_1 belongs to ϵ. To do this we note first that in $\mathfrak{M}_0$, $A - I$ induces a transformation with bounded inverse. Hence if B is equal to this inverse in $\mathfrak{M}_0$ and to zero in $\mathfrak{M}_1$,

$$E_0(A^2 - A)BE_0 = E_0AE_0$$

is in $\mathfrak{I}$, since $A^2 - A$ is. Similarly, it follows that

$$E_1(A - I)E_1 = E_1AE_1 - E_1$$

is in $\mathfrak{I}$. But then, adding the right members of the two preceding equations we find that

$$E_0AE_0 + E_1AE_1 - E_1 = A - E_1$$

is in $\mathcal{J}$, and from this it follows that E_1 belongs to ϵ.

Again the converse part of the theorem is obvious, so the proof is complete.

THEOREM 2.5. *Let $\mathcal{J}$ be an ideal in $\mathcal{B}$, $\mathcal{J}_0$ the set of all nonnegative definite self-adjoint elements of $\mathcal{J}$, $\mathcal{J}_0^2$ the class of all squares of elements of $\mathcal{J}_0$. Then a necessary and sufficient condition that every partially isometric element ω of $\mathcal{B}/\mathcal{J}$ contain a partially isometric transformation W is that $\mathcal{J}_0 = \mathcal{J}_0^2$.*

Again the case $\mathcal{J} = \mathcal{B}$ is trivial, so we assume $\mathcal{J} \subseteq \mathcal{T}$. Let ω be partially isometric, V an element of ω, $V = UB$ its canonical decomposition. Then $V^*V = B^2$ belongs to an idempotent element ϵ of $\mathcal{B}/\mathcal{J}$. Hence, if we identify B^2 with the self-adjoint transformation A which appears in the proof of Theorem 2.4, we can invoke that theorem to establish the existence of a projection E such that $B^2 - E$ is in $\mathcal{J}$. Moreover, an inspection of the proof reveals also that B^2 commutes with E and induces in the range of E a transformation with bounded inverse. In particular, this implies that E has for its range a subspace of the initial set of U and thus that UE is partially isometric, since $EU^*UE = E$.[17] We shall now show that under the condition of the theorem $B - E$ is in $\mathcal{J}$. We note first that since B^2 and E commute, we have $B^2 - E = (B - E)(B + E)$, and since B is nonnegative, $B + E$ induces in $\mathfrak{R}(E)$ a transformation with bounded inverse. Hence, if C is equal to this inverse in $\mathfrak{R}(E)$, and to zero in $\mathfrak{H} \ominus \mathfrak{R}(E)$, we have $(B^2 - E)C = EB - E$, and $EB - E$ is in $\mathcal{J}$. Moreover $(I - E)(B^2 - E) = (I - E)B^2$ is in $\mathcal{J}$. But if $\mathcal{J}$ has the property described in the theorem, $[(I - E)B^2]^{\frac{1}{2}} = (I - E)B$ is in $\mathcal{J}$, and thus

$$(I - E)B + EB - E = B - E$$

is in $\mathcal{J}$ as we wished to show. But then $V - UE = U(B - E)$ is in $\mathcal{J}$ and $W = UE$, which we have already shown to be partially isometric, belongs to ω.

It remains therefore for the converse part of the theorem to be proved. To this end, we suppose that $\mathcal{J}$ contains a nonnegative definite self-adjoint transformation B^2 such that B does not belong to $\mathcal{J}$, and denote by ω the congruence class in $\mathcal{B}/\mathcal{J}$ to which B belongs. Then, since $B^*B = B^2$ belongs to $\mathcal{J}$, we have ω partially isometric by definition. But if V is a partially isometric transformation in ω, $V^*V = E$ must be congruent to B^2 modulo $\mathcal{J}$, which is to say that E is congruent to zero modulo $\mathcal{J}$. But this implies that E has range with finite dimension number and the range of E is the initial set of V. Thus V is in $\mathcal{F}$ and hence in $\mathcal{J}$. However, $V - B$ is in $\mathcal{J}$ since V is in ω, and as B is by assumption not in $\mathcal{J}$, we have a contradiction. Hence, the condition of the theorem is necessary as well as sufficient.

We may note in passing that the ideals (0), $\mathcal{F}$, $\mathcal{T}$, $\mathcal{B}$ all satisfy the condition of Theorem 2.5, but that these are not the only ideals which do so. Consider, for example, the class of all sequences $\{\lambda_n\}$ of nonnegative numbers such that

[17] By [9], Lemma 4.3.2.

$\sum_{n=1}^{\infty} \lambda_n^p$ converges for some p. It is easily seen that this class is an ideal set in the sense of Definition 1.1. and that the corresponding ideal in B has the property of Theorem 2.5.

For the sake of completeness, we state the following obvious theorem:

THEOREM 2.6. *If W is a partially isometric member of $\mathfrak{B}$, the congruence class of W in $\mathfrak{B}/\mathfrak{I}$ is partially isometric in $\mathfrak{B}/\mathfrak{I}$.*

THEOREM 2.7. *Let $\mathfrak{I}$ be an ideal in $\mathfrak{B}$. Then, if ω is a unitary element of $\mathfrak{B}/\mathfrak{I}$, ω contains a maximal partially isometric transformation*[18] *with deficiency-index $(0, n)$ or $(n, 0)$, $n < \aleph_0$.*

Again the case $\mathfrak{I} = \mathfrak{B}$ is trivial, so we assume $\mathfrak{I} \neq \mathfrak{B}$. Let ω be unitary, V a member of ω, $V = UB$ its canonical decomposition. Since ω is also partially isometric, the first part of the proof of Theorem 2.5 applies to yield the following results: there exists a projection E with range in the initial set of U, and which commutes with B, such that $V^*V - E = B^2 - E$ and $EB - E$ are in $\mathfrak{I}$. But since V^*V is congruent to I modulo $\mathfrak{I}$, $I - E$ is in $\mathfrak{I}$ and thus $(I - E)B$ is in $\mathfrak{I}$. Therefore

$$B - E = (I - E)B + EB - E$$

is in $\mathfrak{I}$. But then $V - UE = UB - UE$ is in $\mathfrak{I}$. Hence $W = UE$ is a partially isometric operator which belongs to ω. Moreover, since ω is unitary, $I - WW^*$ is in $\mathfrak{I}$ and hence, since this operator is a projection and belongs to $\mathfrak{I}$, its range must have a finite dimension number. Similarly, $I - W^*W$ has range with a finite dimension number. Thus both the initial and final sets of W have orthogonal complements with finite dimension numbers. Therefore, if W_1 is the contraction of W with domain $\mathfrak{R}(E)$ and X a maximal partially isometric extension of W_1, X has the property required in the theorem and $X - W$ is in $\mathfrak{I}$. Thus X belongs to ω, and the theorem is proved.

It is important to observe that every unitary element of $\mathfrak{B}/\mathfrak{I}$ does not contain a unitary member of $\mathfrak{B}$, except in the trivial cases $\mathfrak{I} = \mathfrak{B}$, $\mathfrak{I} = (0)$.

To prove this, we consider an isometric transformation X with deficiency-index $(0, n)$, $n < \aleph_0$, and the congruence class ω modulo $\mathfrak{I}$, to which X belongs. Then ω is clearly unitary in $\mathfrak{B}/\mathfrak{I}$ provided $\mathfrak{I} \neq (0)$. Now suppose U is a unitary transformation in ω. Then $U - X$ is in $\mathfrak{F}$, and thus in $\mathfrak{I}$, if $\mathfrak{I} \neq \mathfrak{B}$. Hence $I - U^{-1}X$ is in $\mathfrak{I}$ and $U^{-1}X$ also has deficiency-index $(0, n)$. But by a lemma which the author has proved elsewhere, this is possible if and only if $n = 0$.[19]

THEOREM 2.8. *Let $\mathfrak{I}$ be an ideal in $\mathfrak{B}$ different from (0), W a partially isometric operator in $\mathfrak{B}$ with deficiency-index (m, n) $m, n < \aleph_0$. Then the congruence class ω in $\mathfrak{B}/\mathfrak{I}$ to which W belongs is unitary.*

If W has the properties stated then $I - W^*W$ and $I - WW^*$ are projections

[18] We call a partially isometric operator maximal if the isometric transformation which determines it is maximal. Similarly, we shall have occasion to refer to the deficiency-index of a partially isometric operator.

[19] [3], Lemma 4.1.

which belong to $\mathcal{F}$. Thus, since $\mathcal{F} \subseteq \mathcal{I}$ by Theorem 1.7, both of these operators belong to $\mathcal{I}$ and ω is unitary by definition.

We conclude this section with

THEOREM 2.9. *Let $\mathcal{I}$ be an ideal in $\mathfrak{B}$, $\mathcal{I} \neq \mathfrak{B}$. Then the center of $\mathfrak{B}/\mathcal{I}$, that is, the set of all elements of $\mathfrak{B}/\mathcal{I}$ which commute with every element of $\mathfrak{B}/\mathcal{I}$, is the set of all elements $\lambda \cdot 1$, where λ is a complex number.*

It is clear that the center contains the set of all scalar multiples of the identity; hence we need only show that it contains no other members.

We begin by showing that it is sufficient to consider merely the self-adjoint members of the center. For suppose α belongs to the center. Then $\alpha\beta^* - \beta^*\alpha = 0$ for all β in $\mathfrak{B}/\mathcal{I}$ and hence $(\alpha\beta^* - \beta^*\alpha)^* = \beta\alpha^* - \alpha^*\beta = 0$ for all β in $\mathfrak{B}/\mathcal{I}$. Thus α^* belongs to the center, and consequently the self-adjoint elements $\alpha + \alpha^*$ and $i(\alpha - \alpha^*)$ do also. Now suppose $\alpha + \alpha^* = \lambda \cdot 1$, $i(\alpha - \alpha^*) = \mu \cdot 1$. Then, eliminating α^*, we have $\alpha = (\mu + i\lambda)/2i$. Hence we have only to prove that every self-adjoint member of the center is a scalar multiple of the identity.

In terms of operators in $\mathfrak{B}$, this problem reduces to the following: to show that every self-adjoint operator A in $\mathfrak{B}$ such that $AB - BA$ is in $\mathcal{I}$ for all B in $\mathfrak{B}$ is of the form $T + \lambda I$, where T is in $\mathcal{I}$.

We consider first the case $\mathcal{I} = \mathcal{T}$. So we consider a self-adjoint operator A so that $AB - BA$ is totally continuous for all B in $\mathfrak{B}$. If A is not of the form $T + \lambda I$, T in $\mathcal{T}$, the spectrum of A must contain two distinct points, each of which is either a limit point of the spectrum of A or a characteristic value of infinite multiplicity; for, otherwise, the spectrum of A consists solely of isolated characteristic values of finite multiplicity together with one point μ which is either a limit point or a characteristic value of infinite multiplicity, and in this case $A - \mu I$ clearly belongs to $\mathcal{T}$. Hence if $E(\lambda)$ is the resolution of the identity of A, there exist numbers λ_0, λ_1, λ_2, λ_3 in the spectrum of A, $\lambda_0 < \lambda_1 < \lambda_2 < \lambda_3$ such that $E(\lambda_1) - E(\lambda_0)$ and $E(\lambda_3) - E(\lambda_2)$ have ranges $\mathfrak{M}_1$ and $\mathfrak{M}_2$, respectively, which are Hilbert spaces.

Next let us consider the partially isometric operator W with initial set $\mathfrak{M}_1$ and final set $\mathfrak{M}_2$. We then have, for f in $\mathfrak{M}_1$

$$| WAf - AWf | \geqq | AWf | - | WAf |,$$

and thus, since we also have

$$| AWf | \geqq \lambda_2 | Wf | = \lambda_2 | f |, \; | WAf | = | Af | \leqq \lambda_1 | f |$$

we obtain

$$| WAf - AWf | \geqq (\lambda_2 - \lambda_1)f.$$

Hence $(WA - AW)$ induces on $\mathfrak{M}_1$ a transformation with bounded inverse and therefore $(WA - AW)\mathfrak{M}_1$ is a Hilbert space. Consequently, by a lemma previously referred to, $WA - AW$ is not in $\mathcal{T}$. Hence the assumption that A is not of the form $T + \lambda I$, T in $\mathcal{T}$, is untenable, and the theorem is established for $\mathcal{I} = \mathcal{T}$.

Consider now an arbitrary ideal $\mathcal{I}$, $\mathcal{F} \subset \mathcal{I} \subset \mathcal{T}$, $\mathcal{I} \neq \mathcal{F}$, $\mathcal{I} \neq \mathcal{T}$. As before, we consider a self-adjoint operator A such that $AB - BA$ is in $\mathcal{I}$ for all B in $\mathfrak{B}$. Since $\mathcal{I} \subset \mathcal{T}$, it follows from the preceding result that A is of the form $T + \lambda I$, T in $\mathcal{T}$. Thus $TB - BT$ is in $\mathcal{I}$ for all B in $\mathfrak{B}$. Now let $\{\varphi_n\}$ be a complete orthonormal set of characteristic elements of T; $T\varphi_n = \lambda_n \varphi_n$, $n = 1, 2, \cdots$, and let us suppose that T is not in $\mathcal{I}$. Moreover, let us assume that the sequence $\{\varphi_n\}$ is so arranged that λ_1 is different from zero, while the positive terms of $\{| \lambda_n |\}$ are in monotone order. Let $\{\mu_n\}$ be a sequence of positive numbers in the spectral set of $\mathcal{I}$ with $\mu_1 < | \lambda_1 |$; further let $\{\lambda_{n_k}\}$ be an infinite subsequence of $\{\lambda_n\}$ such that $0 < | \lambda_{n_k} | \leqq \mu_k$, $k = 1, 2, \cdots$.[20] Finally, let E be the projection with range $\mathfrak{M}$ determined by the orthonormal set $\{\varphi_{n_k}\}$ and let W be a partially isometric transformation with initial set $\mathfrak{H}$ and final set $\mathfrak{M}$.

Then $WT - TW$ is in $\mathcal{I}$. Hence $WTW^* - TWW^* = WTW^* - TE$ is in $\mathcal{I}$. But then, since $\{\lambda_{n_k}\}$ is in the spectral set of $\mathcal{I}$ by choice of that subsequence, TE is in $\mathcal{I}$. Hence WTW^* is in $\mathcal{I}$, and therefore $W^*WTW^*W = T$ is in $\mathcal{I}$ too, which is a contradiction.

It remains to prove the theorem for the case $\mathcal{I} = \mathcal{F}$. Let us suppose that $BA - AB$ is in $\mathcal{F}$ for all B in $\mathfrak{B}$ and that A is not in $\mathcal{F}$. Then A is of the form $T + \lambda I$, where T is in $\mathcal{T}$ and not in $\mathcal{F}$. Hence there exists an infinite orthonormal set $\{\varphi_n\}$ in $\mathfrak{H}$ such that $T\varphi_n = \lambda_n \varphi_n$, $\lambda_n \neq 0$, $n = 1, 2, \cdots$, $\lambda_n \neq \lambda_m$ if $m \neq n$. Hence, if U is defined by the equations

$$U\varphi_{2n} = \varphi_{2n-1}, \qquad U\varphi_{2n-1} = \varphi_{2n}, \qquad n = 1, 2, \cdots,$$

in the closed linear manifold determined by $\{\varphi_n\}$, $U = I$ in $\mathfrak{H} \ominus \mathfrak{M}$, $UTU^{-1} - T$ is not in $\mathcal{F}$ and hence $UT - TU$ is not either. Therefore the assumption that A is not in $\mathcal{F}$ leads to a contradiction and the theorem is proved for $\mathcal{I} = \mathcal{F}$.

3. A metric in $\mathfrak{B}/\mathcal{T}$

We now confine our attention to the case $\mathcal{I} = \mathcal{T}$, beginning with the definition of a norm in $\mathfrak{B}/\mathcal{T}$.

Throughout the remainder of the paper we employ the notation $| A |$ for the bound of the operator A of $\mathfrak{B}$.

DEFINITION 3.1. *Let α be an arbitrary element of $\mathfrak{B}/\mathcal{T}$. We define $| \alpha |$, called the norm of α, by the equation*

$$| \alpha | = \underset{A \in \alpha}{\text{g. l. b.}} | A |.$$

THEOREM 3.1. *The norm $| \alpha |$ in $\mathfrak{B}/\mathcal{T}$ has the following properties:*

(1) $| \alpha | \geqq 0$, *the equality sign holding if and only if* $\alpha = 0$;

(2) $$| \alpha + \beta | \leqq | \alpha | + | \beta |;$$

(3) $$| \alpha\beta | \leqq | \alpha || \beta |;$$

(4) $$| a\alpha | = | a || \alpha |;$$

(5) $$| \alpha^* | = | \alpha |;$$

(6) $$| 1 | = 1.$$

[20] If no such subsequence exists, T is in $\mathcal{F}$ and hence in $\mathcal{I}$.

The validity of the laws (2) — (5) is an immediate consequence of the definition of $|\alpha|$ in terms of the norm $|A|$ in $\mathfrak{B}$, and the fact that the latter function has those properties. The same is true of the assertion $|\alpha| \geqq 0$. Moreover, in view of (2) and (3) we can conclude that the set of elements α of $\mathfrak{B}/\mathfrak{T}$ for which $|\alpha| = 0$ is a two-sided ideal in $\mathfrak{B}/\mathfrak{T}$. Hence, since $\mathfrak{T}$ is a prime ideal in $\mathfrak{B}$, we have either $\alpha = 0$ when and only when $|\alpha| = 0$, or $\alpha = 0$ for all α in $\mathfrak{B}/\mathfrak{T}$.[21] Thus to complete the proof it is necessary only to establish (6). We have then to show that

$$\operatorname*{g.\,l.\,b.}_{T \epsilon \mathfrak{T}} |I + T| = 1.$$

We note first

$$|I + (T + T^*)/2| \leqq |(I + T)/2| + |(I + T^*)/2| = |I + T|;$$

hence we need only show g. l. b. $|I + T| = 1$ for $T \epsilon \mathfrak{T}$, $T = T^*$. But, if T is so restricted, we can find, for any $\epsilon > 0$, a real number λ with $|\lambda| < \epsilon$ such that $T\varphi = \lambda\varphi$ for some $\varphi \neq 0$ in $\mathfrak{H}$, by virtue of the spectral properties of T. Thus $|(I + T)\varphi| / |\varphi| > (1 - \epsilon)$ and hence $|I + T|$ exceeds $1 - \epsilon$. But then, since ϵ is an arbitrary positive number, we have $|I + T| \geqq 1$. Hence, since $|I| = 1$, (6) follows.

THEOREM 3.2. *With $|\alpha - \beta|$ interpreted as the distance between α and β, $\mathfrak{B}/\mathfrak{T}$ is a complete linear metric space.*

That $\mathfrak{B}/\mathfrak{T}$ is a metric space follows from Theorem 3.1 while its linear properties are evident. To show that it is complete we must prove that, for every sequence $\{\alpha_n\}$ in $\mathfrak{B}/\mathfrak{T}$ such that

$$\lim_{n,m\to\infty} |\alpha_n - \alpha_m| = 0,$$

there exists an element α such that

$$\lim_{n\to\infty} |\alpha_n - \alpha| = 0.$$

Let $\{\alpha_n\}$ be a sequence satisfying the first of these conditions. We choose a subsequence $\{\alpha_{n_k}\}$ such that

$$|\alpha_{n_k} - \alpha_n| \leqq \frac{1}{2^{k+1}}, \qquad \text{for } n \geqq n_k.$$

We then choose an arbitrary element A_{n_1} of α_{n_1} and an element C_1 of $\alpha_{n_2} - \alpha_{n_1}$ such that $|C_1| \leqq \frac{1}{2}$. Setting $A_{n_2} = C_1 + A_{n_1}$, we have A_{n_2} in α_{n_2}

$$|A_{n_2} - A_{n_1}| \leqq \tfrac{1}{2}.$$

[21] For the fact that $\mathfrak{T}$ is divisorless in $\mathfrak{B}$ implies that $\mathfrak{B}/\mathfrak{T}$ contains no two-sided ideals except (0) and $\mathfrak{B}/\mathfrak{T}$ itself. Cf. [19], pp. 56–57 for a discussion of ideals in commutative rings which is readily generalized to cover the case in hand.

Continuing this process we determine a sequence $\{A_{n_k}\}$, with $A_{n_k} \epsilon \alpha_{n_k}$, such that

$$|A_{n_{k+1}} - A_{n_k}| \leqq \frac{1}{2^k}.$$

Thus

$$|A_{n_{k+j}} - A_{n_k}| \leqq \sum_{h=k}^{j-1} |A_{n_{h+1}} - A_{n_h}| \leqq \frac{1}{2^{k-1}}.$$

Hence there exist an element A of $\mathfrak{B}$ such that

$$\lim_{k \to \infty} |A_{n_k} - A| = 0,$$

and if α is the residue class to which A belongs we have, in consequence,

$$\lim_{n \to \infty} |\alpha_n - \alpha| = \lim_{k \to \infty} |\alpha_{n_k} - \alpha| = 0.$$

We note in passing that the space $\mathfrak{B}/\mathfrak{T}$ is non-separable. For every idempotent except 0 in $\mathfrak{B}/\mathfrak{T}$ can be shown to have the norm 1, and if $E(\lambda)$ is the resolution of the identity in $\mathfrak{B}$ of a transformation with spectrum the entire interval $0 \leqq \lambda \leqq 1$, the set $\epsilon(\lambda)$ of idempotents in $\mathfrak{B}/\mathfrak{T}$ such that $\epsilon(\lambda)$ contains $E(\lambda)$ has the property that $\epsilon(\lambda_2) - \epsilon(\lambda_1)$ is an idempotent different from zero if $0 \leqq \lambda_2 < \lambda_1 \leqq 1$.

4. The space $\mathfrak{L}$

We now propose to realize $\mathfrak{B}/\mathfrak{T}$ as an algebraic ring of operators in a certain complex Euclidean space. To define a space $\mathfrak{L}$ suitable for this purpose, we make use of a concept of generalized limit introduced by Banach and Mazur.[22] In the interests of greater generality, however, we shall employ a less restrictive concept of generalized limit than that of these writers, and we begin with a discussion of this concept.[23]

We consider a linear functional defined for all bounded sequences $\{x_n\}$ of real numbers, denoted by $\operatorname{Lim}_{n \to \infty} x_n$, which has the following properties:

(a) $$\operatorname{Lim}_{n \to \infty} (x_n + y_n) = \operatorname{Lim}_{n \to \infty} x_n + \operatorname{Lim}_{n \to \infty} y_n;$$

(b) $$\operatorname{Lim}_{n \to \infty} x_n \geqq 0, \quad \text{for} \quad x_n \geqq 0, \qquad n = 1, 2, \cdots;$$

(c) $\operatorname{Lim}_{n \to \infty} x_n$ is independent of x_p for each integer p;

(d) $$\operatorname{Lim}_{n \to \infty} 1 = 1.$$

---

[22] [2], p. 34.

[23] The possibility of generalizing the notion of Banach in this way was pointed out to us by J. v. Neumann; originally we had employed the Banach limit.

For subsequent use, we note that (a) and (b) imply

(e) $$\operatorname*{Lim}_{n\to\infty} x_n \geqq \operatorname*{Lim}_{n} y_n, \quad \text{if} \quad x_n \geqq y_n \quad \text{for} \quad n = 1, 2, \cdots.$$

The reader will observe that the four preceding conditions differ from the four basic properties of the Banach limit, as given in the reference cited, in the following respects: first, we do not require homogeneity; second, and more important, the Banach limit has the property

(1) $$\operatorname*{Lim}_{n\to\infty} x_{n+1} = \operatorname*{Lim}_{n\to\infty} x_n$$

in place of our (c). Since from (1) one has

$$\operatorname*{Lim}_{n\to\infty} x_{n+p+1} = \operatorname*{Lim}_{n\to\infty} x_n,$$

it is clear that (1) implies (c).[24]

We now wish to show that homogeneity is a consequence of conditions (a) – (d) and that the use of (c) instead of (1) does not affect the other essential properties of $\operatorname*{Lim}_{n\to\infty} x_n$.

To begin we observe that (a) implies

(2) $$\operatorname*{Lim}_{n\to\infty} rx_n = r \operatorname*{Lim}_{n\to\infty} x_n \text{ for all rational } r.$$

We now consider an arbitrary bounded sequence $\{x_n\}$, and rational upper and lower bounds, R and r, respectively, of $\{x_n\}$. Invoking (2), (d), and (e), we then obtain

$$R = \operatorname*{Lim}_{n\to\infty} R \geqq \operatorname*{Lim}_{n\to\infty} x_n \geqq \operatorname*{Lim}_{n\to\infty} r = r.$$

Hence, since R is any rational upper bound of $\{x_n\}$, and r any rational lower bound, we have

(3) $$\text{l. u. b. } x_n \geqq \operatorname*{Lim}_{n\to\infty} x_n \geqq \text{g. l. b. } x_n$$

Moreover, if we now invoke (c) in conjunction with (3), it becomes clear at once that we must have

(f) $$\limsup_{n\to\infty} x_n \geqq \operatorname*{Lim}_{n\to\infty} x_n \geqq \liminf_{n\to\infty} x_n$$

Thus, we have the important property

(g) $$\operatorname*{Lim}_{n\to\infty} x_n = \lim_{n\to\infty} x_n \quad \text{whenever } \lim_{n\to\infty} x_n \text{ exists.}$$

[24] The argument of Banach, loc. cit., thus serves to establish the existence of a functional with the properties (*a*)–(*d*). An elegant and simple direct proof of the existence of such functionals has been obtained by Ulam and Kakutani independently, but has not been published. Moreover, it is not difficult to show that there exist functionals satisfying (*a*)–(*d*) but not (1). This latter fact, however, is not of essential importance in the present paper, but rather in certain related investigations. Added in proof: Since the completion of this paper J. v. Neumann has developed a general theory of limits of the sort used here. His results will appear in a forthcoming number of the ANNALS OF MATHEMATICS STUDIES dealing with the theory of measure.

Furthermore, since it is now clear that $\operatorname*{Lim}_{n\to\infty}$ is a bounded additive functional on the space of all bounded sequences, we can conclude that it is homogeneous:

(h) $$\operatorname*{Lim}_{n\to\infty} ax_n = a \operatorname*{Lim}_{n\to\infty} x_n, \quad \text{for all numbers } a.$$

We now extend the notion of generalized limit to bounded sequences of complex numbers in the obvious way; if $\{x_n\}$ is such a sequence we set

(i) $$\operatorname*{Lim}_{n\to\infty} x_n = \operatorname*{Lim}_{n\to\infty} \mathfrak{R}x_n + i \operatorname*{Lim}_{n\to\infty} \mathfrak{I}x_n .$$

It is then readily proved that properties (a), (c), (d), (g), (h), persist in the complex case.

Hereafter, as occasion requires, we shall refer to the properties of $\operatorname*{Lim}_{n\to\infty} x_n$ as given above by letter without other comment.

We turn now to the construction of our space $\mathfrak{L}$. First, we consider the class $\mathfrak{L}''$ of all sequences $\{f_n\}$ in Hilbert space $\mathfrak{H}$ which are weakly convergent to zero; this class is evidently a module when we define addition and scalar multiplication by the equations

$$\{f_n\} + \{g_n\} = \{f_n + g_n\},$$
$$a\{f_n\} = \{af_n\}.$$

In $\mathfrak{L}''$, we define $(\{f_n\}, \{g_n\})$ by the equation

$$(\{f_n\}, \{g_n\}) = \operatorname*{Lim}_{n\to\infty} (f_n, g_n),$$

invoking the boundedness of the sequences $\{f_n\}$ and $\{g_n\}$ to assure the boundedness of the sequence of numbers $\{(f_n, g_n)\}$. Then from the properties of $\operatorname*{Lim}_{n\to\infty}$ given above and the properties of the inner product in $\mathfrak{H}$, we have

$$(a\{f_n\} + b\{g_n\}, \{h_n\}) = a(\{f_n\}, \{h_n\})b(\{g_n\}, \{h_n\}),$$
$$(\{f_n\}, \{g_n\}) = (\{g_n\}, \{f_n\}),$$
$$(\{f_n\}, \{f_n\}) \geqq 0.$$

Thus $(\{f_n\}, \{g_n\})$ has all the properties requisite for an inner product in $\mathfrak{L}''$ except that requiring that $(\{f_n\}, \{f_n\}) = 0$ if and only if $\{f_n\} = 0$, and it is easy to see that this requirement is not fulfilled, since $\mathfrak{L}''$ contains sequences strongly convergent to zero. However, as A. E. Taylor [18] has pointed out, this requirement is not an essential one, since we can regard it not as a postulate but as a definition of zero and thus of equality. In our case, this means that we must identify the sequences $\{f_n\}$ and $\{g_n\}$ provided

$$\operatorname*{Lim}_{n\to\infty} |f_n - g_n|^2 = 0.$$

In this way, we obtain a class $\mathfrak{L}'$ of elements $f, g, \cdots$, the quotient group of the additive group $\mathfrak{L}''$ by the subgroup of elements $\{f_n\}$ such that

$$\lim_{n\to\infty} |f_n|^2 = 0.$$

If under this homomorphism $\{f_n\} \to f$, $\{g_n\} \to g$, we define

$$(f, g) = (\{f_n\}, \{g_n\}).$$

Thus we achieve in $\mathfrak{L}'$ a (possibly incomplete) complex Euclidean space; that is to say, $\mathfrak{L}'$ is a module and the function (f, g) has the properties

$$(af + bg, h) = a(f, h) + b(g, h),$$
$$(f, g) = \overline{(g, f)},$$
$$(f, f) \geqq 0,$$
$$(f, f) = 0 \quad \text{implies} \quad f = 0.$$

Since the pertinent facts in this connection are discussed in the paper of Taylor cited above, it is unnecessary for us to dwell on them here.

It now remains for us to consider the space $\mathfrak{L}'$ with reference to the matters of completeness and separability. The answers to both questions are provided through the two simple lemmas which follow.

LEMMA 4.1. *The cardinal number of $\mathfrak{L}'$ is not greater than the cardinal number c of the continuum.*

Let $\{\varphi_m\}$ be an arbitrary complete orthonormal set in $\mathfrak{H}$, $\{f_n\}$ an arbitrary element of $\mathfrak{L}''$. Then, if

$$f_n = \sum_{m=1}^{\infty} a_{n,m}\varphi_m$$

$\{a_{n,m}\}$ is a bounded matrix. Thus the cardinal number of $\mathfrak{L}''$ does not exceed the cardinal number of the class of all bounded infinite matrices, and the cardinal number of the latter class is $\mathfrak{c}$. Hence the cardinal number of $\mathfrak{L}'$ is certainly less than $\mathfrak{c}$.

LEMMA 4.2. *The space $\mathfrak{L}'$ contains an orthonormal set with cardinal number* $\mathfrak{c}$.

We consider an enumeration $\{r_n\}$ of the positive rational numbers, a complete orthonormal set $\{\varphi_n\}$ in $\mathfrak{H}$, and the correspondence $r_n \leftrightarrow \varphi_n$ between them. We denote by $\{\{\varphi_n^\alpha\}\}$ the class of all infinite subsequences of $\{\varphi_n\}$, α running over some set which we leave undesignated. Now let a_1 and a_2 be any two distinct positive numbers, $\{r_n'\}$ and $\{r_n''\}$ infinite subsequences of $\{r_n\}$ convergent to a_1 and a_2 respectively. Corresponding to $\{r_n'\}$ and $\{r_n''\}$ we have two subsequences of $\{\varphi_n\}$ which belong to $\{\{\varphi_n^\alpha\}\}$; we denote these by $\{\varphi_n^{\alpha_1}\}$ and $\{\varphi_n^{\alpha_2}\}$, respectively. Then, for n larger than some integer N, we have $\varphi_n^{\alpha_1} \neq \varphi_n^{\alpha_2}$ and hence

$$\lim_{n\to\infty} (\varphi_n^{\alpha_1}, \varphi_n^{\alpha_2}) = 0.$$

Moreover,

$$\operatorname*{Lim}_{n\to\infty} | \varphi_n^\alpha |^2 = 1,$$

for all α. Hence, if $\boldsymbol{\varphi}_\alpha$ denotes the element of $\mathfrak{L}'$ containing $\{\varphi_n^\alpha\}$, it follows that the set of all $\{\boldsymbol{\varphi}_\alpha\}$ contains an orthonormal set with cardinal number c.

THEOREM 4.1. *The space $\mathfrak{L}'$ is incomplete.*

Consider an orthonormal set $\{\psi_\alpha\}$ in $\mathfrak{L}'$ with cardinal number at least $\mathfrak{c}$. Then the set of all $\sum_\alpha a_\alpha \psi_\alpha$ with $\sum_\alpha | a_\alpha |^2 < \alpha$ has cardinal number at least $2^{\mathfrak{c}}$, and hence, in view of Lemma 4.1, this set cannot belong to $\mathfrak{L}'$. Therefore $\mathfrak{L}'$ is incomplete.[25]

We now denote by $\mathfrak{L}$ the space obtained by completing $\mathfrak{L}'$; the details of this construction are described in [5] and in [13], so we need not consider them here.

THEOREM 4.2. *The dimension number of $\mathfrak{L}$ is $\mathfrak{c}$.*

Since $\mathfrak{L}'$ is dense in $\mathfrak{L}$, the dimension number of $\mathfrak{L}$ cannot exceed $\mathfrak{c}$, by Lemma 4.1. But by Lemma 4.2, it cannot be less than $\mathfrak{c}$.

5. The algebraic ring $\mathfrak{M}$ and congruence modulo $\mathcal{T}$ in $\mathcal{B}$

We now consider transformations induced in the space $\mathfrak{L}$ by means of members of the ring $\mathcal{B}$.

LEMMA 5.1. *Let A be an arbitrary bounded everywhere-defined transformation in $\mathfrak{H}$, $\{f_n\}$ an arbitrary sequence of the class $\mathfrak{L}''$. Then $\{Af_n\}$ is in $\mathfrak{L}''$ and* $\operatorname*{Lim}_{n\to\infty} | Af_n |^2 = 0$ *if* $\operatorname*{Lim}_{n\to\infty} | f_n |^2 = 0$.

That $\{Af_n\}$ is in $\mathfrak{L}''$ follows at once from the fact that a bounded transformation is weakly continuous. And since we have $| Af_n |^2 \leqq | A |^2 | f_n |^2$, $n = 1, 2, \cdots$, it follows from property (e) of $\operatorname*{Lim}_{n\to\infty}$ that $\operatorname*{Lim}_{n\to\infty} | f_n |^2 = 0$ implies

$$\operatorname*{Lim}_{n\to\infty} | Af_n |^2 = 0.$$

THEOREM 5.1. *Let A be an arbitrary member of $\mathcal{B}$. Then, if f is an arbitrary element of $\mathfrak{L}'$, and $\{f_n\}$ belongs to f, we set*

$$g = T_1(A)f,$$

where g is the element of $\mathfrak{L}'$ containing $\{Af_n\}$. The transformation $T_1(A)$ so defined in $\mathfrak{L}'$ is a single valued linear bounded transformation with bound not exceeding the bound of A in $\mathfrak{H}$. Thus $T_1(A)$ has a unique closed bounded extension $T(A)$ with domain $\mathfrak{L}$, and the bound of $T(A)$ does not exceed the bound of A in $\mathfrak{H}$.

That $T_1(A)$ is single-valued follows at once from Lemma 5.1, while its linear character is a consequence of the linearity of A. Since, in addition,

$$| T_1(A)f |^2 = \operatorname*{Lim}_{n\to\infty} |Af_n|^2 \leqq \operatorname*{Lim}_{n\to\infty} | A |^2 | f_n |^2 = | A |^2 | f |^2$$

[25] This simple proof of Theorem 4.1 was suggested by J. v. Neumann. The theorem can also be proved directly.

it is evident that $T_1(A)$ is bounded with bound less than or equal to $|A|$. Thus the transformation $\tilde{T}_1(A) = T(A)$ exists and has domain $\mathfrak{L}$, while its bound is clearly the same as that of $T_1(A)$.

THEOREM 5.2. *The class $\mathfrak{M}$ of operators $T(A)$ in $\mathfrak{L}$, defined for all A in $\mathfrak{B}$, is an algebraic ring of operators in the class of all bounded everywhere defined operators in $\mathfrak{L}$, and is a homomorphism of $\mathfrak{B}$ with respect to the operations $+, \cdot, ^*$, and scalar multiplication; that is,*

$$T(A + B) = T(A) + T(B), \qquad T(AB) = T(A)T(B),$$
$$T(A^*) = T^*(A), \qquad T(aA) = aT(A).$$

All of these relations are quite obvious, except possibly $T(A^*) = T^*(A)$. To prove this we consider two arbitrary elements f and g of $\mathfrak{L}'$ and sequences $\{f_n\}$ and $\{g_n\}$ belonging to f and g, respectively. Then

$$(T(A)f, g) = \operatorname*{Lim}_{n\to\infty} (Af_n, g_n) = \operatorname*{Lim}_{n\to\infty} (f_n, A^*g_n) = (f, T(A^*)g).$$

Thus $T(A^*)$ and $T^*(A)$ coincide on $\mathfrak{L}'$, and therefore throughout $\mathfrak{L}$.

LEMMA 5.2. *A necessary and sufficient condition that $T(A)$ be the transformation in $\mathfrak{L}$ which takes every element of $\mathfrak{L}$ into zero is that A belong to the ideal $\mathfrak{T}$ of totally continuous operators in $\mathfrak{B}$. Thus $T(A) = T(B)$ if and only if A is congruent to B modulo $\mathfrak{T}$.*

From the homomorphism $\mathfrak{B} \to \mathfrak{M}$, it follows that the set of all A in $\mathfrak{B}$ such that $T(A) = 0$ is a two-sided ideal $\mathfrak{I}$ in $\mathfrak{B}$. Moreover, since a totally continuous transformation A in $\mathfrak{B}$ takes weakly convergent sequences into strongly convergent ones, it follows from property (e) of $\operatorname*{Lim}_{n\to\infty}$ that $\mathfrak{I} \supseteqq \mathfrak{T}$. Hence, by Theorem 1.4, we have either $\mathfrak{I} = \mathfrak{T}$ or $\mathfrak{I} = \mathfrak{B}$. But since $\mathfrak{I}$ clearly fails to contain the identity in $\mathfrak{B}$, we must conclude that $\mathfrak{I} = \mathfrak{T}$, which establishes the lemma.

Since it also follows immediately from the homomorphism $\mathfrak{B} \to \mathfrak{M}$ that $\mathfrak{M}$ is isomorphic to the ring $\mathfrak{B}/\mathfrak{I}$, where $\mathfrak{I}$ is the ideal of all A in $\mathfrak{B}$ such that $T(A) = 0$, we can now conclude that $\mathfrak{M}$ is isomorphic to $\mathfrak{B}/\mathfrak{T}$. More precisely, we have

THEOREM 5.3. *The algebraic ring $\mathfrak{M}$ is isomorphic to $\mathfrak{B}/\mathfrak{T}$ with respect to the operations $+, \cdot, ^*, a\cdot$, an element $T(A)$ of $\mathfrak{M}$ corresponding to the element α of $\mathfrak{B}/\mathfrak{T}$ if and only if A belongs to α.*

DEFINITION 5.1. *If α is an arbitrary element of $\mathfrak{B}/\mathfrak{T}$, we define $T(\alpha)$ as the element of $\mathfrak{M}$ corresponding to α under the isomorphism of Theorem 5.3.*

Evidently $T(\alpha)$ is identical with $T(A)$, for all A in α, and we shall continue to use both notations for elements of $\mathfrak{M}$ as occasion requires.

THEOREM 5.4. *An element $T(\alpha)$ of $\mathfrak{M}$ is self-adjoint, partially isometric, or unitary, respectively if and only if α has that property in the sense of Definition 2.1. An element $T(\alpha)$ of $\mathfrak{M}$ is a projection if and only if α is an idempotent according to that definition.*

The assertion of the theorem concerning self-adjointness is obvious. To prove the other parts of the theorem we note first that since the properties which form

the various criteria of Definition 2.1 are all defined in terms of the operations $\cdot$ and $*$, it follows from Theorem 5.3 that an element α of $\mathfrak{B}/\mathfrak{T}$ possesses one of them if and only if $T(\alpha)$ does. But for transformations, each of these properties is characteristic of the class of transformations in question: more precisely, an everywhere defined bounded linear operator T in $\mathfrak{L}$ is a projection if and only if $T^2 = T^* = T$,[26] is partially isometric if and only if T^*T is a projection,[27] and thus is obviously unitary if and only if T^*T and TT^* are equal to the identity in $\mathfrak{L}$. Hence the theorem follows.

The reader will note that Theorems 2.3–2.7 can be interpreted now to yield assertions concerning the homomorphism $\mathfrak{B} \rightarrow \mathfrak{M}$; the details here are obvious and we omit them.

THEOREM 5.5. *Let α be an arbitrary element of $\mathfrak{B}/\mathfrak{T}$. Then the bound of the operator $T(\alpha)$ in $\mathfrak{L}$ is $|\alpha|$; in other words, the isomorphism $\mathfrak{B}/\mathfrak{T} \leftrightarrow \mathfrak{M}$ is an isometry. Thus $\mathfrak{M}$ is closed in the uniform topology for operators.*

Evidently the concluding assertion is a consequence of the first one and Theorem 3.2. Hence we need only show $|T(\alpha)| = |\alpha|$.

From the final statement of Theorem 5.1 and the definition of the norm in $\mathfrak{B}/\mathfrak{T}$, we have at once

$$|T(\alpha)| \leqq |\alpha|. \tag{1}$$

Hence we need only establish

$$|T(\alpha)| \geqq |\alpha|. \tag{2}$$

To prove (2), we first select an arbitrary element A of α, with canonical decomposition $A = WB$. We then denote by λ the lowest upper bound of those points of the spectrum of B which are either limit points of the spectrum or characteristic values of infinite multiplicity, and by S the set of points μ in the spectrum of B such that μ exceeds λ. Then S clearly consists entirely of isolated points, each a characteristic value of finite multiplicity. Furthermore, either S is a finite sequence $\{\mu_n\}$ or an infinite sequence with λ as limit. Hence if $\mathfrak{M}_n$ is the characteristic manifold of B corresponding to μ_n, $n = 1, 2, \cdots$, and we set $C = B - \lambda I$ on $\mathfrak{M} = \sum_n \mathfrak{M}_n$, $C = 0$ on $\mathfrak{H} \ominus \mathfrak{M}$, C belongs to $\mathfrak{T}$. Thus, if $B_1 = B - C$, then $A_1 = WB_1$ belongs to α. Moreover, this is evidently the canonical decomposition of A_1, so $|A_1| = |B_1|$. Hence we have

$$|B_1| \geqq |\alpha|. \tag{3}$$

Now let us consider the transformation $T(B_1)$ in $\mathfrak{L}$. We distinguish two cases, according as the sequence $\{\mu_n\}$ is infinite or finite. If $\{\mu_n\}$ is infinite, $\mathfrak{M} = \sum_n \mathfrak{M}_n$ is a Hilbert space and contains an infinite orthonormal set $\{\varphi_n\}$. Furthermore, $B_1\varphi_n = \lambda\varphi_n$, $n = 1, 2, \cdots$. Thus, if φ is the element of $\mathfrak{L}'$ to which

[26] [17], Theorems 2.35, 2.36.

[27] [9], Lemma 4.3.2.

$\{\varphi_n\}$ belongs, we have

$$(4)\qquad \begin{aligned}(T(B_1)\boldsymbol{\varphi}, \boldsymbol{\varphi}) &= \operatorname*{Lim}_{n\to\infty} (B_1\varphi_n, \varphi_n) = \lambda.\\ (\boldsymbol{\varphi}, \boldsymbol{\varphi}) &= \operatorname*{Lim}_{n\to\infty} (\varphi_n, \varphi_n) = 1.\end{aligned}$$

But clearly $\lambda = | B_1 |$, and hence we have

$$(5)\qquad | T(B_1) | \geqq | B_1 |.$$

Now suppose $\{\mu_n\}$ is finite, and let $E(\lambda)$ be the resolution of the identity for B_1 in $\mathfrak{H}$. Then, since λ is a limit point of the spectrum of B_1, there exists a monotone increasing sequence $\{\lambda_n\}$ with limit λ, such that $E(\lambda_{n+1}) - E(\lambda_n)$ is different from zero, $n = 1, 2, \cdots$. Hence we can select an orthonormal set $\{\varphi_n\}$ with φ_n in the range of $E(\lambda_{n+1}) - E(\lambda_n)$, $n = 1, 2, \cdots$. Moreover, for every n, we have $\lambda_n \leqq (B_1\varphi_n, \varphi_n) \leqq \lambda_{n+1}$. Hence if $\boldsymbol{\varphi}$ is the element of $\mathfrak{L}'$ containing $\{\varphi_n\}$, we have (4) in this case also.

Finally, since $B_1 = W^*A_1$ we have $T(B_1) = T(W^*)T(A_1)$ and since $T(W^*)$ is partially isometric by Theorems 5.4 and 2.6, it follows that we have $| T(W^*)| = 1$, and hence that

$$(6)\qquad | T(A_1)| \geqq | T(B_1)|$$

holds. But $T(A_1) = T(\alpha)$, and hence from (3), (5) and (6) we have (2) which completes the proof of the theorem.

We now wish to prove that $\mathfrak{M}$ is not closed in the weak or strong topologies for operators. The proof reposes on the following lemma concerning monotone sequences of projections in $\mathfrak{M}$.

LEMMA 5.3. *Let $\{T(\epsilon_n)\}$ be a sequence of projections in $\mathfrak{M}$ such that $T(\epsilon_{n+1}) \leqq T(\epsilon_n)$, $n = 1, 2, \cdots$. Then, if $\lim_{n\to\infty} T(\epsilon_n) = 0$, $T(\epsilon_n) = 0$ for all n greater than some integer M.*

By Theorem 5.4, each of the terms of $\{\epsilon_n\}$ is an idempotent, and consequently by Theorem 2.4, contains a projection E_n of $\mathfrak{B}$. The sequence $\{E_n\}$, however, is evidently not necessarily monotone, and our next step is to show that there exists a monotone non-increasing sequence $\{F_n\}$ of projections in $\mathfrak{B}$ such that $E_n - F_n$ is in $\mathfrak{T}$, $n = 1, 2, \cdots$. We begin by setting $F_1 = E_1$ and then, assuming that F_n is determined for all $n \leqq N$, we show that F_{N+1} can be defined.

We note first that

$$T(\epsilon_n)T(\epsilon_{n+1})T(\epsilon_n) = T(\epsilon_{n+1})$$

and hence that $F_NE_{N+1}F_N - E_{N+1}$ is in $\mathfrak{T}$. Consequently, it follows that $(F_NE_{N+1}F_N)^2 - F_NE_{N+1}F_N$ is in $\mathfrak{T}$. Thus, if $\mathfrak{M}_N$ is the range of F_N, $F_NE_{N+1}F_N$ induces in $\mathfrak{M}_N$ a self-adjoint transformation E'_{N+1} congruent to its square modulo the class of all totally continuous operators in $\mathfrak{M}_N$. Hence, by Theorem 2.4, there exists in $\mathfrak{M}_N$ a projection F'_{N+1} congruent to E'_{N+1} modulo that class. Therefore if F_{N+1} is the projection in $\mathfrak{H}$ which is equal to F'_{N+1} in $\mathfrak{M}_N$, equal to

zero in $\mathfrak{H} \ominus \mathfrak{M}_N$, then $F_{N+1} - F_N E_{N+1} F_N$ is in $\mathfrak{T}$. But then $F_{N+1} - E_{N+1}$ is in $\mathfrak{J}$ and since F_{N+1} is equal to zero in $\mathfrak{H} \ominus \mathfrak{M}_N$, we have $F_{N+1} \leqq F_N$. Thus the sequence $\{F_n\}$ with the stated properties exists.

Now let us suppose that $T(\epsilon_n)$ is never zero. Then, if $\lim_{n\to\infty} T(\epsilon_n) = 0$, $T(\epsilon_n)$ must be different from $T(\epsilon_{n+1})$ for an infinite number of values of n. Hence we can select a subsequence $\{T(\epsilon_{n_k})\}$ such that $T(\epsilon_{n_{k+1}}) < T(\epsilon_{n_k})$ holds, $k = 1, 2, \cdots$. Consequently, we clearly have $F_{n_{k+1}} < F_{n_k}$, $k = 1, 2, \cdots$, since $T(\epsilon_n) = T(F_n)$. Therefore, if $\mathfrak{N}_k$ is the range of F_{n_k}, none of the spaces $\mathfrak{N}_k \ominus \mathfrak{N}_{k+1}$ is empty, and we can select an orthonormal set $\{\varphi_k\}$ with φ_k in $\mathfrak{N}_k \ominus \mathfrak{N}_{k+1}$, $k = 1, 2, \cdots$. Let $\mathfrak{N}_0$ be the closed linear manifold determined by $\{\varphi_k\}$, F_0 the projection with range $\mathfrak{N}_0$. Then $F_0 - F_0 F_{n_k} F_0$ is the projection with range determined by $\{\varphi_j\}$, $j = 1, 2, \cdots k$, and hence belongs to $\mathfrak{T}$. Consequently $T(F_0)T(F_{n_k})T(F_0) = T(F_0)$ and $T(F_0)$ is clearly not zero. Hence, since we obviously have

$$\lim_{k\to\infty} T(F_0)T(F_{n_k})T(F_0) = 0$$

if $\lim_{n\to\infty} T(\epsilon_n) = 0$, it follows that the latter is impossible. Therefore $T(\epsilon_n) = 0$ for all n sufficiently large, as we wished to show.

It is of some interest to note the following alternative statement of the preceding lemma; if $\{T(\epsilon_n)\}$ is an infinite sequence of orthogonal projections in $\mathfrak{M}$, and if $\sum_{n=1}^{\infty} T(\epsilon_n)$ is the identity in $\mathfrak{K}$, then all but a finite number of terms of the sequence $T(\epsilon_n)$ are zero.

THEOREM 5.6. *The algebraic ring $\mathfrak{M}$ is not closed in either the weak or strong topology for operators.*

We consider a monotone sequence $\{T(\epsilon_n)\}$ of projections in $\mathfrak{M}$, with $T(\epsilon_{n+1}) > T(\epsilon_n)$, $n = 1, 2, \cdots$. Such a sequence is readily generated by means of a sequence $\{E_n\}$ of projections in $\mathfrak{B}$ such that $E_{n+1} - E_n$ is a projection with range a Hilbert space, $n = 1, 2, \cdots$; we have only to choose ϵ_n as the residue class of E_n. The sequence $T(\epsilon_n)$, being monotone, is convergent in both the weak and the strong topologies and has a limit which is a projection. Let us suppose that this limit is in $\mathfrak{M}$ and hence that it has the form $T(\epsilon)$ where ϵ is an idempotent in $\mathfrak{B}/\mathfrak{T}$. Then $\{T(\epsilon_n) - T(\epsilon)\}$ is a monotone non-increasing sequence of projections in $\mathfrak{M}$ with limit zero and thus $T(\epsilon_n) = T(\epsilon)$ for all n greater than some integer N, by Lemma 5.3. This, however, is impossible in view of the fact that $\{T(\epsilon_n)\}$ was chosen with $T(\epsilon_{n+1}) > T(\epsilon_n)$, $n = 1, 2, \cdots$; we must conclude therefore that the sequence $\{T(\epsilon_n)\}$ has no limit in $\mathfrak{M}$.

We proceed now to examine the relationship between the spectrum of a self-adjoint member of $\mathfrak{M}$ and the corresponding self-adjoint transformations in $\mathfrak{B}$. We begin with a formal definition.

DEFINITION 5.2. *Let A be an arbitrary self-adjoint transformation in $\mathfrak{H}$. Then a point λ of the spectrum of A which is a limit point of the spectrum or a character-*

istic value of infinite multiplicity is called a point of condensation of the spectrum.[28] *The set of all such points is called the condensed spectrum of* A. *The complementary set in the* λ-*plane is called the augmented resolvent set of* A.

THEOREM 5.7. *Let* A *be a self-adjoint transformation in* $\mathfrak{H}$. *Then a necessary and sufficient condition that* λ *belong to the augmented resolvent set of* A *is that the manifold* $\mathfrak{M}$ *of solutions of the equation* $Af - \lambda f = 0$ *have a finite dimension number and that in* $\mathfrak{H} \ominus \mathfrak{M}$, $A - \lambda I$ *induce a transformation with bounded inverse.*

Let $\mathfrak{M}$ be the manifold of zeros of A, A_1 the transformation induced in $\mathfrak{H} \ominus \mathfrak{M}$ by A. Then the condition of the theorem may be stated in this way: $\mathfrak{M}$ is finite-dimensional and λ is in the resolvent set of A_1. But the latter is possible if and only if λ is a finite distance from the spectrum of A_1, which is to say that λ is an isolated point of the spectrum of A or belongs to the resolvent set of A. Thus the theorem follows.

THEOREM 5.8. *Let* A *be a self-adjoint transformation in* $\mathfrak{H}$. *Then the resolvent set of the transformation* $T(A)$ *in* $\mathfrak{L}$ *is the augmented resolvent set of* A *and the spectrum of* $T(A)$ *is the condensed spectrum of* A. *Every point in the spectrum of* $T(A)$ *is a characteristic value with multiplicity* $\mathfrak{c}$.

Let λ belong to the augmented resolvent set of A, and let $\mathfrak{M}$ be the manifold of zeros of $A - \lambda I$, E the projection operator of $\mathfrak{H} \ominus \mathfrak{M}$. Let B be equal in $\mathfrak{M}$ to zero, and in $\mathfrak{H} \ominus \mathfrak{M}$ to the inverse of the transformation induced in that space by $A - \lambda I$. Then B is in $\mathfrak{B}$ and $B(A - \lambda I) = E$. Thus $T(B)T(A - \lambda I) = T(E)$. But $T(E)$ is the identity in $\mathfrak{L}$, since $I - E$ is in $\mathfrak{T}$, and hence $T(B) = [T(A) - \lambda T(I)]^{-1}$. Therefore λ is in the resolvent set of $T(A)$.

Now suppose λ is in the condensed spectrum of A. Then λ is either a characteristic value of infinite multiplicity or a limit point of the spectrum of A, and in either case we can select an orthonormal set $\{\varphi_n\}$ in $\mathfrak{H}$ such that

$$\lim_{n \to \infty} (A\varphi_n, \varphi_n) = \lambda$$

$$\lim_{n \to \infty} (A\varphi_n, A\varphi_n) = \lambda^2.$$

Thus

$$\lim_{n \to \infty} | A\varphi_n - \lambda\varphi_n |^2 = \lim_{n \to \infty} (| A\varphi_n |^2 + \lambda^2 | \varphi_n |^2 - 2\lambda(A\varphi_n, \varphi_n)) = 0.$$

But then, if $\{\psi_n\}$ is any subsequence of $\{\varphi_n\}$ and $\boldsymbol{\psi}$ the element of $\mathfrak{L}'$ containing $\{\psi_n\}$, we have

$$T(A)\boldsymbol{\psi} = \lambda\boldsymbol{\psi}.$$

Moreover, there exist $\mathfrak{c}$ such subsequences such that any two have only a finite number of terms in common and consequently $\mathfrak{c}$ orthogonal elements of $\mathfrak{L}'$ satisfying the preceding equation. Therefore λ is a characteristic value of $T(A)$ with multiplicity $\mathfrak{c}$.

---

[28] These are the Häufungspunkte of the spectrum in the sense of Weyl, [20].

We have now shown that every point of the augmented resolvent set of A is in the resolvent set of $T(A)$ and that every point of the condensed spectrum of A is in the spectrum of $T(A)$. But since the augmented resolvent set and the condensed spectrum of A together constitute the entire λ-plane, they must be respectively the entire resolvent set and the entire spectrum of A.

It should be observed that we cannot infer from theorem 5.8 that the sum of the characteristic manifolds of $T(A)$ is $\mathfrak{L}$. Whether or not this is so we are unable to say at present.

We now have from Theorem 5.8 the classical theorem of Weyl [20].

THEOREM 5.9. *Let A and B be two self-adjoint transformations in $\mathfrak{H}$, such that $A - B$ is totally continuous. Then A and B have the same condensed spectrum and the same augmented resolvent set.*[29]

The theorem is obvious since $T(A) = T(B)$.

We proceed now to determine the resolution of the identity corresponding to a self-adjoint transformation in $\mathfrak{M}$ by means of the resolution of the identity of a corresponding member of $\mathfrak{B}$. If A is self-adjoint in $\mathfrak{H}$, $E(\lambda)$ its resolution of the identity, it is clear that $T(E(\lambda))$ has many of the properties of a resolution of the identity in $\mathfrak{L}$. Specifically, it is readily shown that the following assertions hold:

(1) $T(E(\lambda))$ permutes with $T(A)$;

(2) $T(E(\lambda))T(E(\mu)) = T(E(\mu))\, T(E(\lambda)) = T(E(\mu))$ for $\mu \leqq \lambda$;

(3) $T(E(\lambda)) = 0$ if λ is less than the lower bound of $T(A)$, $T(E(\lambda)) = 1$ if λ exceeds the upper bound of $T(A)$;

(4) in the range of $T(E(\lambda))$, the upper bound of $T(A)$ does not exceed λ, and in the range of $1 - T(E(\lambda))$ the lower bound of $T(A)$ is not less than λ.

In spite of these facts $T(E(\lambda))$ fails usually to be the resolution of the identity for $T(A)$. This is readily seen in view of Lemma 5.3, which assures us that in general we do not have

$$\lim_{\epsilon \to 0} T(E(\lambda + \epsilon)) = T(E(\lambda)).$$

Moreover, if A and B are self-adjoint and congruent modulo $\mathfrak{T}$, $E(\lambda)$ and $F(\lambda)$ their respective resolutions of the identity, we do not in general have $E(\lambda) - F(\lambda)$ in $\mathfrak{T}$. Thus for a self-adjoint transformation $T(A)$ in $\mathfrak{M}$ we can exhibit many monotone families of projections with the properties (1) – (4) above, none of which is the resolution of the identity of A.

We can however, derive from any one of these families of projections the resolution of the identity belonging to the member of $\mathfrak{M}$ in question. The procedure is described in the following theorem.

THEOREM 5.10. *Let A be a self-adjoint transformation in $\mathfrak{H}$, $E(\lambda)$ its resolution of the identity. Let $\boldsymbol{E}(\lambda)$ be the family of projections in $\mathfrak{L}$ defined by the equation*

$$\boldsymbol{E}(\lambda) = \lim_{\epsilon \to 0} T(E(\lambda + \epsilon)), \qquad \epsilon > 0.$$

[29] Weyl proves also that if A is an arbitrary bounded self-adjoint transformation, there exists a totally continuous self-adjoint transformation T such that $A + T$ has a pure point spectrum. At present we see no direct way to derive this result from ours. Cf. also [14].

Then $\boldsymbol{E}(\lambda)$ *is the resolution of the identity in* $\mathfrak{L}$ *of* $T(A)$.

The existence of the limit $\boldsymbol{E}(\lambda)$ follows of course from the monotone character of the family $T(E(\lambda))$ which follows in turn from the corresponding property of $E(\lambda)$. We shall now show that $\boldsymbol{E}(\lambda)$ has the following six properties:

(1) $\boldsymbol{E}(\lambda)$ permutes with $T(A)$;

(2) $\boldsymbol{E}(\lambda)E(\mu) = \boldsymbol{E}(\mu)\boldsymbol{E}(\lambda) = \boldsymbol{E}(\mu)$ for $\mu \leqq \lambda$;

(3) $\boldsymbol{E}(\lambda) = 0$ for $\lambda < a$ (a the lower bound of $T(A)$), and $\boldsymbol{E}(\lambda) = 1$ for $\lambda \geqq b$ (b the upper bound of $T(A)$);

(4) $\lim_{\epsilon \to 0} \boldsymbol{E}(\lambda + \epsilon) = \boldsymbol{E}(\lambda)$, $\epsilon > 0$;

(5) in the range of $\boldsymbol{E}(\lambda)$, the upper bound of $T(A)$ does not exceed λ;

(6) in the range of $1 - \boldsymbol{E}(\lambda)$, the lower bound of $T(A)$ is not less than λ and if it is equal to λ it is not attained.

The validity of (1) follows at once from the permutability of $T(A)$ and $T(E(\lambda))$.

To prove (2), we note first that

$$T(E(\lambda + \epsilon))T(E(\mu)) = T(E(\mu))T(E(\lambda + \epsilon)) \quad \text{for all} \quad \epsilon > 0.$$

Thus, allowing ϵ to tend to zero, we have

$$\boldsymbol{E}(\lambda)T(E(\mu)) = T(E(\mu))\boldsymbol{E}(\lambda).$$

Hence

$$\boldsymbol{E}(\lambda)T(E(\mu + \epsilon)) = T(E(\mu + \epsilon))\boldsymbol{E}(\lambda)$$

for all μ and all $\epsilon > 0$. Consequently, again allowing ϵ to approach zero, we have

$$\boldsymbol{E}(\lambda)\boldsymbol{E}(\mu) = \boldsymbol{E}(\mu)\boldsymbol{E}(\lambda).$$

Moreover, if $\mu \leqq \lambda$, we have clearly $\boldsymbol{E}(\mu) \leqq \boldsymbol{E}(\lambda)$, and so $\boldsymbol{E}(\lambda)\boldsymbol{E}(\mu) = \boldsymbol{E}(\mu)$. Therefore (2) holds.

Now let a be the lower bound of $T(A)$. Then by Theorem 5.8, a is the lower bound of the condensed spectrum of A and it follows therefore that if λ is less than a, then the range of $E(\lambda)$ must be finite-dimensional. For otherwise the spectrum of A would have points of condensation less than or equal to λ. Thus $T(E(\lambda)) = 0$ for $\lambda < a$ and therefore $\boldsymbol{E}(\lambda) = 0$ for $\lambda < a$.

On the other hand, let b be the upper bound of the spectrum of $T(A)$. Then b is the upper bound of the condensed spectrum of A and by a similar argument $I - E(\lambda)$ has a finite-dimensional range, $\lambda \geqq b$. Thus $T(E(\lambda + \epsilon)) = T(I)$ for $\lambda \geqq b$ and $\epsilon > 0$. But then $\boldsymbol{E}(\lambda)$ is the identity in $\mathfrak{L}$ for $\lambda \geqq b$. Thus we have (3).

Next we note that for $\epsilon_1 > \epsilon > 0$, we have

$$\boldsymbol{E}(\lambda) \leqq \boldsymbol{E}(\lambda + \epsilon) \leqq T(E(\lambda + \epsilon_1)).$$

Hence, allowing ϵ and ϵ_1 both to approach zero, preserving the relations $\epsilon_1 > \epsilon$, $\epsilon_1 > 0$, we obtain (4).

Finally, we consider the behavior of $T(A)$ in the range of $\boldsymbol{E}(\lambda)$ and its orthog-

onal complement. Since $E(\lambda)AE(\lambda)$ has upper bound not exceeding λ, it follows that the upper bound of $T(E(\lambda + \epsilon))T(A)T(E(\lambda + \epsilon))$ does not exceed $\lambda + \epsilon$. But for every f in $\mathfrak{L}$, we have

$$(\boldsymbol{E}(\lambda)T(A)\boldsymbol{E}(\lambda)f, f) = \lim_{\epsilon \to 0} (T(E(\lambda + \epsilon))T(A)T(E(\lambda + \epsilon))f, f), \qquad \epsilon > 0.$$

Hence the upper bound of $\boldsymbol{E}(\lambda)T(A)\boldsymbol{E}(\lambda)$ does not exceed λ and (5) is established.

In entirely similar fashion, it can be shown that in the range of $1 - \boldsymbol{E}(\lambda)$ the lower bound of $T(A)$ is not less than λ. Moreover, if λ is the lower bound and this lower bound is attained, we have for some f in $\mathfrak{L}$,

$$(1 - \boldsymbol{E}(\lambda))T(A)(1 - \boldsymbol{E}(\lambda))f = \lambda f.$$ [30]

Hence

$$\lim_{\epsilon \to 0} (T(A)(1 - \boldsymbol{E}(\lambda + \epsilon))f, (1 - \boldsymbol{E}(\lambda + \epsilon))f) = \lambda \, | f |^2.$$

But $(T(A)\ (1 - \boldsymbol{E}(\lambda + \epsilon))f,\ (1 - \boldsymbol{E}(\lambda + \epsilon))f)$ is monotone non-decreasing as ϵ approaches zero, and consequently we have

$$(T(A)(1 - \boldsymbol{E}(\lambda + \epsilon))f, (1 - \boldsymbol{E}(\lambda + \epsilon))f) \leqq \lambda \, | f |^2$$

for some $\epsilon > 0$. This, however is impossible, since in the range of $1 - \boldsymbol{E}(\lambda + \epsilon)$, the lower bound of $T(A)$ is not less than λ. Hence, we have (6).

These six properties are sufficient to characterize $\boldsymbol{E}(\lambda)$ as the resolution of the identity of $T(A)$. One can for example, argue as follows: On the basis of these six properties the approximation theorem (Theorem 6) of the paper [6] of Lengyel and Stone can be proved and from this result it follows that $\boldsymbol{E}(\lambda)$ permutes with every bounded linear operator defined over $\mathfrak{L}$ which permutes with $T(A)$ ([6], Theorem 7). But this fact together with properties (1), (5), (6) above uniquely determines $\boldsymbol{E}(\lambda)$ as the resolution of the identity for $T(A)$ ([6], Theorem 5).

We are now in position to establish certain relationships between the resolutions of the identity belonging to two self-adjoint transformations in $\mathfrak{H}$ whose difference is totally continuous.

THEOREM 5.11. *Let A and B be self-adjoint transformations in $\mathfrak{H}$ such that $A - B$ is totally continuous. Let $E(\lambda)$ and $F(\lambda)$ be the resolutions of the identity corresponding to A and B respectively. Then, if μ is in the augumented resolvent set of A, $E(\mu) - F(\mu)$ is totally continuous.*

Let $\boldsymbol{E}(\lambda)$ be the resolution of the identity of $T(A)$ in $\mathfrak{L}$. Then, since μ is in the resolvent set of $T(A)$, we have for some $\delta > 0$, $\boldsymbol{E}(\mu - \delta) = \boldsymbol{E}(\mu + \delta)$. But, from Theorem 5.10 it is clear that we have

$$\boldsymbol{E}(\mu - \delta) \leqq T(E(\mu)) \leqq \boldsymbol{E}(\mu + \delta),$$

$$\boldsymbol{E}(\mu - \delta) \leqq T(F(\mu)) \leqq \boldsymbol{E}(\mu + \delta).$$

[30] [6], Theorem 3, for example.

Thus $T(E(\mu)) = T(F(\mu))$ and hence $E(\mu) - F(\mu)$ belongs to $\mathfrak{T}$, as we wished to show.

It is important to note that the requirement that μ be in the augmented resolvent set cannot be dropped from Theorem 5.11. Consider, for example, a nonnegative definite self-adjoint transformation A in $\mathfrak{H}$ which is totally continuous and has an inverse, and its resolution of the identity $E(\lambda)$. Consider also the transformation 0 in $\mathfrak{H}$, and its resolution of the identity $F(\lambda)$. Then $A - 0$ is in $\mathfrak{T}$, but $E(0) = 0$ and $F(0) = I$.

THEOREM 5.12.[31] *Let A and B be self-adjoint transformations in $\mathfrak{H}$ such that $A - B$ is totally continuous, and let $E(\lambda)$ and $F(\lambda)$ be their respective resolutions of the identity. Then, if μ is less than λ, there exists a totally continuous transformation $T_{\lambda,\mu}$, in B such that $E(\mu) + T_{\mu,\lambda}$ is a projection satisfying the inequality*

$$E(\mu) + T_{\mu,\lambda} \leqq F(\lambda).$$

If $T_{\mu,\lambda}$ can be chosen so that the equality sign holds, then the only points λ_0 on the interval $\mu < \lambda_0 < \lambda$ which belong to the spectrum of either A or B are characteristic values of finite multiplicity.

Conversely, let A and B be self-adjoint transformations in $\mathfrak{H}$ with resolutions of the identity $E(\lambda)$ and $F(\lambda)$, respectively. Then, if the inequalities $T(E(\mu)) \leqq T(F(\lambda))$ and $T(F(\mu)) \leqq T(E(\lambda))$ hold for all λ and μ such that $\mu < \lambda$, $A - B$ is totally continuous.

We prove the converse part first. From the inequalities in question, it follows at once that

$$\lim_{\epsilon \to 0} T(E(\lambda + \epsilon)) = \lim_{\epsilon \to 0} T(F(\lambda + \epsilon)), \qquad \epsilon > 0,$$

for all λ and thus that $T(A)$ and $T(B)$ have the same resolution of the identity. Hence $T(A) = T(B)$, or $A - B$ is in $\mathfrak{T}$.

Now let $A - B$ belong to $\mathfrak{T}$, and let $\boldsymbol{E}(\lambda)$ be the resolution of the identity of $T(A)$ in $\mathfrak{L}$. Then, for $\mu < \lambda$, we have

$$T(E(\mu)) \leqq \boldsymbol{E}(\mu) \leqq T(F(\lambda)).$$

Hence $T(F(\lambda))T(E(\mu))T(F(\lambda)) = T(E(\mu))$ and therefore $F(\lambda)E(\mu)F(\lambda) - E(\mu)$ is in $\mathfrak{T}$. Consequently, invoking Theorem 2.4 with reference to the ring of bounded everywhere defined operators in the range of $F(\lambda)$, we see that there exists a totally continuous operator T_0 in $\mathfrak{B}$, with the value zero in the range of $I - F(\lambda)$, such that

$$F(\lambda)E(\mu)F(\lambda) + T_0$$

is a projection satisfying

$$F(\lambda)E(\mu)F(\lambda) + T_0 \leqq F(\lambda).$$

[31] Compare §4 of [20] to which this theorem is closely related.

Thus, if we set

$$T_{\mu,\lambda} = T_0 + F(\lambda)E(\mu)F(\lambda) - E(\mu),$$

$T_{\mu,\lambda}$, is in $\mathfrak{T}$ and $E(\mu) + T_{\mu,\lambda}$ is a projection satisfying the inequality of the theorem.

Now suppose we have

$$E_\mu + T_{\mu,\lambda} = F(\lambda).$$

Then $T(E(\mu)) = T(F(\lambda))$, and since for all λ_0 on the interval $\mu < \lambda_0 < \lambda$, we have

$$T(E(\mu)) \leqq E(\lambda_0) \leqq T(F(\lambda)),$$

it follows that $E(\lambda)$ is constant on that interval. Hence every point on the interval is in the resolvent set of $T(A) = T(B)$, and this is equivalent to the concluding statement of the first paragraph of the theorem.

COROLLARY. *Let A, B, $E(\lambda)$, and $F(\lambda)$ be as in Theorem 5.12. Then, if $\lambda \neq \mu$,*

$$E(\mu)F(\lambda) - F(\lambda)E(\mu)$$

is totally continuous.

From the inequality of Theorem 5.12, we have, for $\mu < \lambda$,

$$F(\lambda)E(\mu) + F(\lambda)T_{\mu,\lambda} = E(\mu)F(\lambda) + T_{\mu,\lambda}F(\lambda),$$

or

$$F(\lambda)E(\mu) - E(\mu)F(\lambda) = T_{\mu,\lambda}F(\lambda) - F(\lambda)T_{\mu,\lambda}.$$

Thus $F(\lambda)E(\mu) - E(\mu)F(\lambda)$ is totally continuous for $\mu < \lambda$, and by symmetry for $\mu > \lambda$ also.

This is not necessarily so, however, for $\lambda = \mu$. For consider any two projections E and F in $\mathfrak{H}$ such that $\mathfrak{R}(E)$ and $\mathfrak{R}(F)$ and their orthogonal complements are Hilbert spaces. Let A be a transformation which is equal to zero in $\mathfrak{R}(E)$ and which induces in $\mathfrak{H} \ominus \mathfrak{R}(E)$ a nonnegative definite self-adjoint totally continuous transformation whose inverse exists. Then A is in $\mathfrak{T}$ and if $E(\lambda)$ is the resolution of the identity of A, $E(0) = E$. Similarly, we can define a self-adjoint totally continuous operator B with resolution of the identity $F(\lambda)$ such that $F(0) = F$. But then $A - B$ is in $\mathfrak{T}$, while in general $E(0)F(0) - F(0)E(0)$ is not.

It follows also from the preceding example that the hypothesis $\mu < \lambda$ of Theorem 5.12 cannot be replaced by the hypothesis $\mu \leqq \lambda$.

We conclude this paper with a few observations concerning congruence modulo $\mathfrak{T}$ in $\mathfrak{B}$ which are not restricted to self-adjoint transformations.

THEOREM 5.13. *Let A_1 and A_2 be members of $\mathfrak{B}$ whose difference belongs to $\mathfrak{T}$, $A_1 = W_1B_1$, $A_2 = W_2B_2$ their canonical decompositions. Then $B_1 - B_2$ is in $\mathfrak{T}$, and if zero is in the augmented resolvent set of B_1, $W_1 - W_2$ is in $\mathfrak{T}$.*

Since $A_1^*A_1 = B_1^2$, $A_2^*A_2 = B_2^2$, we have $T(A_1^*A_1) = T(B_1^2) = T(B_2^2)$. More-

over, $[T(B_1)]^2 = [T(B_2)]^2 = T(B_1^2)$. But B_1 and B_2 are nonnegative definite and hence

$$\operatorname*{Lim}_{n\to\infty} (B_1 f_n, f_n) \geqq 0, \qquad \operatorname*{Lim}_{n\to\infty} (B_2 f_n, f_n) \geqq 0$$

for all $\{f_n\}$ in $\mathfrak{L}''$, by property (*f*) of $\operatorname*{Lim}_{n\to\infty}$. Hence $T(B_1)$ and $T(B_2)$ are both nonnegative definite, and since $T(B_1^2)$ can have only one nonnegative definite square root, we have $T(B_1) = T(B_2)$ which implies that $B_1 - B_2$ is in $\mathcal{T}$.

Now let

$$T(A_1) = \boldsymbol{W}\boldsymbol{B}$$

be the canonical decomposition of $T(A_1)$. Then $\boldsymbol{B} = T(B_1) = T(B_2)$. Moreover,

$$T(A_1) = T(W_1)T(B_1) = T(W_2)T(B_2).$$

Hence we can infer that the initial set of the partially isometric transformation $T(W_1)$ contains the initial set of $\boldsymbol{W}$ (the closure of the range of $\boldsymbol{B}$) and that the two transformations are equal there. Hence $T(W_1)$ is identical with $\boldsymbol{W}$ provided the initial set of $T(W_1)$ is the closure of the range of $\boldsymbol{B}$. But, if the origin is in the augmented resolvent set of B_1, it is in the resolvent set of $T(B_1) = \boldsymbol{B}$ and hence $\boldsymbol{W}$ is unitary. Thus under the hypothesis of the theorem we have $T(W_1) = \boldsymbol{W}$. Furthermore, under that hypothesis, the origin is also in the augmented resolvent set of B_2 and hence a similar argument yields the equation $T(W_2) = \boldsymbol{W}$. Thus $T(W_1) = T(W_2)$ and $W_1 - W_2$ is in $\mathcal{T}$ as we wished to show.

We wish to emphasize that the hypothesis that zero be in the augmented resolvent set of B_1 cannot be omitted from Theorem 5.13. For example, consider a totally continuous nonnegative definite self-adjoint transformation A with an inverse. Then if $A = W_1B_1$ is its canonical decomposition and $0 = W_2B_2$ is the canonical decomposition of 0, we have $W_1 = I$, $W_2 = 0$, $A - 0$ in $\mathcal{T}$.

We now extend Theorem 5.9 to cover transformations which are not necessarily self-adjoint.

DEFINITION 5.3. *Let A be an arbitrary bounded everywhere defined transformation in $\mathfrak{H}$. We define the augmented resolvent set of A as the set of points λ in the complex plane such that the following conditions are satisfied:*
(1) *the manifold of zeros of $A - \lambda I$ has a finite dimension number;*
(2) *the range of $A - \lambda I$ is closed;*
(3) *the orthogonal complement of the range of $A - \lambda I$ has a finite dimension number.*
The complement of the augmented resolvent set is called the condensed spectrum of A.

Evidently the augmented resolvent set of A contains the resolvent set, so that our terminology is justified.

THEOREM 5.14. *Let A and C be two bounded everywhere-defined transforma-*

tions in $\mathfrak{H}$ such that $A - C$ is totally continuous. Then A and C have the same augmented resolvent sets.

It is sufficient for the proof to show that every point in the augmented resolvent set of A is also in the augmented resolvent set of C.

Let λ be a point of the augmented resolvent set of A, $\mathfrak{M}$ the manifold of zeros of $A - \lambda I$. Then on $\mathfrak{H} \ominus \mathfrak{M}$, $A - \lambda I$ induces a transformation T with an inverse whose range is $\mathfrak{R}(A - \lambda I)$. Moreover, since $\mathfrak{R}(A - \lambda I)$ is closed, T^{-1} is bounded.

Now let $\mathfrak{N}$ be the manifold of zeros of $C - \lambda I$. Then if $\mathfrak{N}$ is a Hilbert space, $(\mathfrak{H} \ominus \mathfrak{M})\cdot\mathfrak{N}$ is a Hilbert space, and T has a contraction T_1 with domain $(\mathfrak{H} \ominus \mathfrak{M})\cdot\mathfrak{N}$ which also has a bounded inverse. But in $(\mathfrak{H} \ominus \mathfrak{M})\cdot\mathfrak{N}$, we have $A - C = T_1$ and this contradicts our hypothesis that $A - C$ is in $\mathfrak{T}$. Hence $\mathfrak{N}$ must have a finite dimension number.

We now observe that when λ is a point of the augmented resolvent set of A, $\bar{\lambda}$ is a point of the augmented resolvent set of A^*. Hence the preceding argument serves to show that the manifold of zeros of $C^* - \bar{\lambda} I$ has a finite dimension number. But this manifold is precisely the orthogonal complement of the closure of $\mathfrak{R}(C - \lambda I)$, and hence to show that λ is in the augmented resolvent set of C, it remains only for us to prove that $\mathfrak{R}(C - \lambda I)$ is closed.

To establish the latter we consider the canonical decompositions

$$A - \lambda I = W_1 B_1, \qquad C - \lambda I = W_2 B_2 .$$

From Theorem 5.13, it follows that $B_1 - B_2$ is totally continuous. Moreover, since the range of $A - \lambda I$ is closed, so also is the range of B_1, while the manifold of zeros of B_1 is the manifold of zeros of $A - \lambda I$. But then the origin is in the augmented resolvent set of B_1 and hence in the augmented resolvent set of B_2. Hence $\mathfrak{R}(B_2)$ is closed which implies that $\mathfrak{R}(C - \lambda I)$ is closed also as we wished to show.

If λ is in the augmented resolvent set of A, we denote by R_λ the transformation which is equal to zero in $\mathfrak{H} \ominus \mathfrak{R}(A - \lambda I)$ and which takes each element f of $\mathfrak{R}(A - \lambda I)$ into that element g in the orthogonal complement of the manifold of zeros of $(A - \lambda I)$ which satisfies $(A - \lambda I)g = f$. Thus $R_\lambda(A - \lambda I)$ is the projection with range the orthogonal complement of the manifold of zeros of $A - \lambda I$. We call the family of transformations R_λ so defined the augmented resolvent of A.

THEOREM 5.15. *Let A and B be two transformations in $\mathfrak{H}$ such that $A - B$ is in $\mathfrak{T}$. Let $R_\lambda^{(1)}$ and $R_\lambda^{(2)}$ be respectively their augmented resolvents. Then $R_\lambda^{(1)} - R_\lambda^{(2)}$ is in $\mathfrak{T}$ for all λ for which these transformations are defined.*

Consider the transformation $T(A) = T(B)$ in $\mathfrak{L}$. If λ is in the augmented resolvent set of A (or B) it is in the resolvent set of $T(A)$, and $T(R_\lambda^{(1)})$ and $T(R_\lambda^{(2)})$ are both inverses of $T(A) - \lambda\cdot 1 = T(B) - \lambda\cdot 1$. But only one such inverse can exist; hence $T(R_\lambda^{(1)}) = T(R_\lambda^{(2)})$ and $R_\lambda^{(1)} - R_\lambda^{(2)}$ is in $\mathfrak{T}$.

ILLINOIS INSTITUTE OF TECHNOLOGY

REFERENCES

1. A. A. ALBERT, *Modern Higher Algebra*, Chicago, 1937.
2. S. BANACH, *Théorie des opérations linéaires*, Warsaw, 1932.
3. J. W. CALKIN, *Abstract Symmetric Boundary Conditions*, Transactions of the American Mathematical Society, 45 (1939), pp. 369-442.
4. R. COURANT, *Zur Theorie der linearen Integralgleichungen*, Mathematische Annalen, 89 (1923), pp. 161-178.
5. K. FRIEDRICHS, *Spektraltheorie halbbeschrankten Operatoren*, Mathematischen Annalen, 109 (1934), pp. 465-487.
6. B. A. LENGYEL and M. H. STONE, *Elementary Proof of the Spectral Theorem*, Annals of Mathematics, 37 (1936), pp. 853-864.
7. H. LÖWIG, *Komplexe euclidische Räume von beliebiger endlicher oder transfiniter Dimensionzahl*, Acta Litterarum ac Scientiarum, Sectio Scientiarum Mathematicarum, Szeged, 7 (1934), pp. 1-33.
8. F. J. MURRAY, *Linear Transformations between Hilbert Spaces and the Applications of this Theory to Linear Partial Differential Equations*, Transactions of the American Mathematical Society, 37 (1935), pp. 301-338.
9. F. J. MURRAY and J. VON NEUMANN, *On Rings of Operators*, I, Annals of Mathematics, 37 (1936), pp. 116-229.
10. J. VON NEUMANN, *Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren*, Mathematische Annalen, 102 (1929), pp. 49-131.
11. J. VON NEUMANN, *Zur Algebra der Funktionaloperatoren*, ibid., 102 (1929), pp. 370-427.
12. J. VON NEUMANN, *Uber adjungtierte Funktionaloperatoren*, Annals of Mathematics, 33 (1932), pp. 294-310.
13. J. VON NEUMANN, *Lectures on Operator Theory* (mimeographed), Princeton, 1934.
14. J. VON NEUMANN, *Charakterisierung des Spektrums eines Integraloperators*, Actualités Scientifiques et Industrielles 229, Paris, 1935.
15. J. VON NEUMANN, *On Rings of Operators*, III, Annals of Mathematics, 41 (1940), pp. 94-161.
16. F. RELLICH, *Spektraltheorie in nichtseparablen Räumen*, Mathematische Annalen, 110 (1935), pp. 342-356.
17. M. H. STONE, *Linear Transformations in Hilbert Space and their Applications to Analysis*, New York, 1932.
18. A. E. TAYLOR, *The Hilbert Space Postulates—A Further Reduction*, Bulletin of the American Mathematical Society, 41 (1935), pp. 847-848.
19. B. L. VAN DER WAERDEN, *Moderne Algebra*, I, Berlin, 1930.
20. H. WEYL, *Uber beschränkte quadratische Formen, deren Differenz vollstetig ist*, Rendiconti del Circolo Matematico Palermo, 27 (1909), pp. 373-392.

Annals of Mathematics
Vol. 42, No. 4, October, 1941

MEASURE-PRESERVING HOMEOMORPHISMS AND METRICAL TRANSITIVITY*

By J. C. Oxtoby and S. M. Ulam

(Received May 1, 1941)

Introduction

In the study of dynamical systems one is led naturally to the consideration of measure-preserving transformations. A Hamiltonian system of $2n$ differential equations induces in the phase space of the system a measure-preserving flow, that is, a one-parameter group of transformations that leave invariant the $2n$-dimensional measure. Making use of one or more integrals of the system, one obtains a reduced phase manifold of lower dimension which likewise undergoes a flow into itself, and in general admits an invariant measure related in a simple manner to that in the larger phase space. If the differential equations are sufficiently regular the flow will have corresponding properties of continuity and differentiability. Thus the study of one-parameter continuous groups of measure-preserving automorphisms[1] of finite dimensional spaces has an immediate bearing on dynamics and the theory of differential equations.

In statistical mechanics one is especially interested in time-average properties of a system. In the classical theory the assumption was made that the average time spent in any region of phase space is proportional to the volume of the region in terms of the invariant measure, more generally, that time-averages may be replaced by space-averages. To justify this interchange, a number of hypotheses were proposed, variously known as ergodic or quasi-ergodic hypotheses, but a rigorous discussion of the precise conditions under which the interchange is permissible was only made possible in 1931 by the ergodic theorem of Birkhoff.[2] This established the *existence* of the time-averages in question, for almost all initial conditions, and showed that if we neglect sets of measure zero, the interchange of time- and space-averages is permissible if and only if the flow in the phase space is *metrically transitive.* A transformation or a flow is said to be metrically transitive if there do not exist two disjoint invariant sets both having

* This paper includes results presented to the Amer. Math. Soc. in preliminary reports on April 15 and December 30, 1938. Most of the results were obtained while the authors were members of the Harvard Society of Fellows.

[1] An automorphism is a 1:1 bicontinuous transformation of a space onto itself. It is measure-preserving with respect to a measure μ if $\mu TA = \mu A = \mu T^{-1}A$ whenever A is measurable.

[2] G. D. Birkhoff, *Proof of the ergodic theorem.* Proc. Nat. Acad. USA. **17** (1931) 650–660. An interesting connection between this theorem and the fundamental theorem of the calculus has been shown by N. Wiener. See *The ergodic theorem.* Duke Jour. **5** (1939) 1–18.

positive measure.[3] Thus the effect of the ergodic theorem was to replace the ergodic hypothesis by the hypothesis of metrical transitivity.[4]

Nevertheless, in spite of the simplification introduced by the ergodic theorem, the problem of deciding whether particular systems are metrically transitive or not has proved to be very difficult. Hedlund[5] showed that the flow defined by the system of geodesics on certain surfaces of constant negative curvature is metrically transitive, and this result was generalized by Hopf.[6] More recently, the result has been extended to surfaces of variable negative curvature, independently by both Hedlund[7] and Hopf.[8] This important class of systems is so far the only large class known to be metrically transitive. In fact, the only other known examples of metrically transitive continuous flows are the spiral motions on the n-dimensional torus, including the rotation of the circumference of a circle as the simplest example of all. The ergodic properties of these systems were established as long ago as 1916 by Weyl,[9] in a quite different connection.

Thus the known examples of metrically transitive continuous flows are all in manifolds, indeed in manifolds of restricted topological type, either toruses or manifolds of direction elements over surfaces of negative curvature.[10] An outstanding problem in ergodic theory has been the existence question—can a metrically transitive continuous flow exist in an arbitrary manifold, or in any space that is not a manifold? In the present paper we shall obtain a complete answer to this question, at least on the topological level, for polyhedra of dimension three or more. It will appear that the only condition that needs to be imposed is a trivially necessary kind of connectedness. In particular, there exists a metrically transitive continuous flow in the cube, in the solid torus, and in any pseudo-manifold of dimension at least three. Since the phase spaces of dynamical systems have the required kind of connectedness, it follows that the hypothesis of metrical transitivity in dynamics involves no *topological* contradiction. More precisely, in any phase space there can exist a continuous flow

[3] The notion was first introduced in a different connection by G. D. Birkhoff and P. A. Smith. See *Structure analysis of surface transformations*. Liouville's Jour. **7** (1928) 365.

[4] For a historical survey, see G. D. Birkhoff and B. O. Koopman, *Recent contributions to the ergodic theory*. Proc. Nat. Acad. USA. **18** (1932) 279.

[5] G. A. Hedlund, *On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature*. Ann. of Math. **36** (1935) 787. See also his paper, *A new proof for a metrically transitive system*. Amer. J. Math. **62** (1940) 233–242.

[6] E. Hopf, *Fuchsian groups and ergodic theory*. Trans. Amer. Math. Soc. **39** (1936) 299. See also his booklet, *Ergodentheorie*. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. **5**, Berlin 1937.

[7] G. A. Hedlund, Bull. Amer. Math. Soc. Abstract 46–3–173.

[8] E. Hopf, *Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung*. Ber. Verh. Sächs. Akad. Wiss. Leipzig **91** (1939) 261–304.

[9] H. Weyl, *Über die Gleichverteilung von Zahlen mod. Eins*. Math. Ann. **77** (1916) 315.

[10] Some examples of discontinuous flows have been studied. See E. Hopf, *Ergodentheorie*, and J. von Neumann, *Zur Operatorenmethode in der klassischen Mechanik*. Annals of Math. **33** (1932) 587–642.

metrically transitive with respect to the invariant measure associated with the system.

It may be recalled that the original ergodic hypothesis of Boltzmann—that a single streamline passes through all points of phase space—had to be abandoned because it involved a topological impossibility.[10a] It was replaced by a quasi-ergodic hypothesis—that some streamline passes arbitrarily close to all points of phase space. But it is not obvious that even this weak hypothesis is topologically reasonable in general phase spaces, and in any case it is not sufficient to justify the interchange of time- and space-averages. It is therefore of some interest to know that the ergodic hypothesis in its modern form of metrical transitivity is at least free from any objection on topological grounds.

It must be emphasized, however, that our investigation is on the topological level. The flows we construct are continuous groups of measure-preserving automorphisms, but not necessarily differentiable or derivable from differential equations. Thus they correspond to dynamical systems only in a generalized sense. In this respect the flows studied by Hedlund and Hopf are closer to the dynamical problem. In any case, the problem is illuminated by considering it in a more general setting, and probably a certain amount of generalization is necessary in order to give meaning to some questions of the type with which we deal.

The conjecture has frequently been expressed, first by Birkhoff, then by Hopf and others, that metrical transitivity is probably the "general case." The conjecture was never given a precise formulation, but was based on the fact that the transitive case is the non-integrable case, in the sense that no uniform measurable integrals exist, and on the idea that a general transformation or flow should be expected to shuffle points in a more or less random fashion, and therefore leave invariant as few sets as possible. To see the precise connection between metrical transitivity and randomness, let us recall the statement of the ergodic theorem for transformations. It asserts that if μ is a completely additive finite measure in E, and T a measure-preserving transformation of E onto itself, then for any μ-integrable function $f(p)$ on E the average value $\lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} f(T^{\nu}p) = f^*(p)$ exists for almost all points p, is a measurable function of p, and that $\int f^*(p)\, d\mu(p) = \int f(p)\, d\mu(p)$. Metrical transitivity is precisely the condition that $f^*(p)$ should be constant for almost all p. Thus if T is metrically transitive and f_A is the characteristic function of any measurable set A, we have $\frac{\mu A}{\mu E} = \lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} f_A(T^{\nu}p)$ for almost all p. The left member of this equation represents the average number of images of p that fall in A, or the *frequency* with which the images of p fall in the set A under iteration of T, and the equation

[10a] A. Rosenthal, *Beweis der Unmöglichkeit ergodischer Gassysteme*, Ann. der Physik, (4) **42** (1913) 796–806.

asserts that this frequency is proportional to the measure of A. A sequence of points selected at random from E would be expected to have just such a distribution, in fact, the ratio $\mu A/\mu E$ may be interpreted as the probability that a point selected at random lies in A. Thus a metrically transitive transformation is one under which almost all points generate sequences that are distributed like random sequences in respect to the average number that fall in any measurable set. This perhaps makes it seem plausible that a measure-preserving transformation "selected at random" should be metrically transitive.

One might try to make precise the idea that metrical transitivity is the general case by introducing a *measure* in the space of all measure-preserving transformations of E, but it seems difficult to do this in any natural way. Nevertheless, there is a simple and natural *metric* in the space of automorphisms of E, in case E is compact, and one may ask whether in this space metrical transitivity is the general case in the *topological sense*, that is, whether such automorphisms constitute all but a set of first category.[11] This is what we are going to show, under suitable assumptions about E. Thereby we shall dispose of the existence problem at the same time.[12] The result is perhaps a little surprising in view of the fact that metrical transitivity is a purely measure-theoretic property, and it often happens that what is the general case in the sense of measure is exceptional in the sense of category, and vice versa.[13]

The fact that metrical transitivity turns out to be the general case in the topological sense raises the question how far residual sets of measure-preserving automorphisms may exhibit other random properties. For instance, a random sequence is not only distributed so that the average number of points in a set is proportional to its measure, but the limiting frequency is approached in the manner of "Laplace-Liapounoff." Presumably metrical transitivity is not sufficient to insure that almost all points should generate sequences exhibiting this more precise behavior. It would therefore be of interest to know whether transformations of this sort are likewise general, or indeed whether they can exist

---

[11] A set is said to be of first category if it can be represented as a sum of countably many nowhere dense sets. Any other set is said to be of second category. Complements of first category sets are called residual sets.

[12] Baire's Theorem asserts that in a complete metric space every residual set is of second category and is dense. It summarizes concisely a typical form of existence proof.

[13] For instance, the law of large numbers is false in the sense of category. That is, the set of numbers x in $0 < x < 1$ such that in their infinite dyadic development the number of ones in the first n places divided by n tends to one-half is of first category (although of measure one). If x_n denoted the n-th digit in the dyadic development of x, the set in question is represented by

$$\prod_k \sum_N \left[\prod_n \mathop{\mathsf{E}}_x \left\{\left|\frac{x_1 + x_2 + \cdots + x_{N+n}}{N+n} - \frac{1}{2}\right| < \frac{1}{k}\right\}\right],$$

and for $N > 0$ and $k > 2$ the set enclosed in square brackets is nowhere dense, as may be seen by inserting a sufficiently long block of zeros far out in the development of a number.

at all. A result in this direction is given in §12. In the same order of ideas one may ask for automorphisms exhibiting the various types of *mixture* properties.[14]

In the course of our investigation we shall derive a number of results about transformations and measures which have an interest entirely apart from questions relating to transitivity. In particular, the results obtained in Part II concerning measures topologically equivalent to Lebesgue measure appear to be fundamental to the theory of measure-preserving homeomorphisms. In sections 13 to 15 a number of questions of group-theoretic interest are answered on the basis of these results. The paper involves an intimate combination of the methods of topology and measure theory. A suggestive example of the way in which such a combination of methods may lead to purely topological results not otherwise apparent is the sort of topological ergodic theorem obtained in §16. Another result of purely topological character is the Corollary to Lemma 16 concerning equivalence of Cantor sets under automorphism of the containing space. This result is of independent interest since it adds to the results obtained by Antoine[15] relating to this question.

A fundamental outstanding problem in topology is that of approximating to an arbitrary homeomorphism by a differentiable one.[16] In §18, as a by-product of our investigation, we obtain the result that any measure-preserving automorphism of the r-dimensional cube that leaves the boundary fixed can be approximated uniformly by one that is differentiable *almost everywhere*, in fact, it is locally linear about almost all points. This result in itself is not strong enough to have important topological implications, but it suggests that the special properties of measure-preserving automorphisms may enable one to obtain approximation theorems for them that are difficult or perhaps impossible to establish in general. In any case, the very precise properties of the approximating transformation obtained in Theorem 12 may well serve as a basis for answering other questions concerning approximation by automorphisms with special measure-theoretic properties.

I. Preliminary Results

1. Definitions and Principal Theorem

We assume the standard notions[17] of *polyhedron, euclidean polyhedron,* and *complex*, except that we shall always understand these to be finite. Thus a polyhedron is compact. Some of the results can be generalized to infinite polyhedra, but it seems best to leave such extensions out of the present paper.

[14] See G. A. Hedlund, *The dynamics of geodesic flows*. Bull. Amer. Math. Soc. **45** (1939) 241–260.

[15] L. Antoine, *Sur l'homéomorphie de deux figures et de leurs voisinage*. Liouville's Jour. (8) **4** (1921) 221–325 esp. p. 307 ff.

[16] J. W. Alexander, *Some problems in topology*. Verh. des Int. Math. Kong. Zurich 1932, vol. 1, p. 249–257.

[17] For all topological definitions see P. Alexandroff and H. Hopf, *Topologie I*, Berlin 1935, hereafter referred to as AH.

A *regular point* of an r-dimensional polyhedron is a point that has a neighborhood homeomorphic to an open sphere in r-space, any other point of the polyhedron is called *singular*. A polyhedron is called *regularly connected* if its regular points form a connected set dense in the polyhedron.[18]

A *finite outer measure* in a space E is a function μ^* defined for all subsets of E and satisfying the three conditions:

M1: $0 \leqq \mu^*A \leqq \mu^*E, \quad o < \mu^*E < +\infty, \quad \mu^*(\text{void set}) = 0$

M2: $\mu^*A \leqq \mu^*B \quad \text{if } A \subset B$

M3: $\mu^* \sum_{n=1}^{\infty} A_n \leqq \sum_{n=1}^{\infty} \mu^* A_n$.

A set A is *measurable* with respect to μ^* if $\mu^*W = \mu^*WA + \mu^*W(E - A)$ for every set W. The function μ^* is completely additive with respect to its class of measurable sets, and the measure function thus derived from μ^* by restricting its domain is denoted by μ.[19]

A *Carathéodory outer measure* is one that satisfies also the condition

M4: $\mu^*(A + B) = \mu^*A + \mu^*B$ if A and B are separated by a positive distance. The measurable sets then include all Borel sets.

By a *Lebesgue-Stieltjes outer measure* in a polyhedron, we shall understand a finite Carathéodory outer measure that satisfies the further condition

M5: $\mu^*A = \inf_{G \supset A} \mu^*G \qquad G$ open.

A measure derived from such an outer measure will be called a Lebesgue-Stieltjes measure (LS measure).[20]

We introduce the following special definition. A Lebesgue-Stieltjes measure in an r-dimensional polyhedron E, $r \geqq 1$, will be called *r-dimensional* if it is zero for points, zero for the set of singular points, and positive for neighborhoods of regular points. This definition is consistent with ordinary usage, since r-dimensional Lebesgue measure in any r-dimensional euclidean polyhedron evidently fulfills the requirements. The invariant measures associated with dynamical systems are usually defined by integrating a positive density function and are therefore r-dimensional in the present sense.

The set of all automorphisms of a polyhedron E (or of any compact space) is made into a metric space $H[E]$ by the definition

$$\rho(g, h) = \max [\max_{x \in E} \rho_E(gx, hx), \max_{x \in E} \rho_E(g^{-1}x, h^{-1}x)],$$

[18] This is equivalent to the ordinary definition. See AH pp. 400, 402.

[19] For the theory of measure, see Carathéodory, *Vorlesungen über reellen Funktionen*, Leipzig and Berlin 1918; or H. Hahn, *Theorie der reelen Funktionen I*, Berlin 1921.

[20] For polyhedra embedded in euclidean spaces, these measures are the same as those obtained by relativizing measures in the containing space that are derived from non-negative additive functions of an interval. The reasoning is essentially contained in S. Saks, *Theory of the integral*, Warsaw-Lwów 1937, Chap. 3. Thus our definition is consistent with the ordinary concept of a non-negative Lebesgue-Stieltjes measure, and has the advantage of being intrinsic. Also it is equally applicable to curved polyhedra, in which intervals have no meaning. An outer measure that satisfies conditions M1 to M5 is called by Hahn a (finite) Inhaltsfunktion. See Hahn, op. cit. p. 444.

where g, h are any two automorphisms of E, and ρ_E denotes distance in E. In this metric, convergence of a sequence of automorphisms means uniform convergence together with uniform convergence of the sequence of inverses. One can verify without difficulty that the space $H[E]$ is complete and that the group operations (composition and inverse of automorphisms) are continuous in this metric, so that $H[E]$ is a *metric group*, that is, a space of type (G) in the sense of (e.g.) Banach.[21] The distance of an automorphism from the identity will be called its *norm*, it is equal to the maximum distance through which any point is displaced by the automorphism. An *arbitrarily small* automorphism is one whose norm is arbitrarily small.

It may be remarked that whereas uniform convergence of a sequence of automorphisms h_n does not insure the existence of a limiting automorphism, nevertheless uniform convergence of h_n *to an automorphism* h implies uniform convergence of h_n^{-1} to h^{-1}, therefore convergence in $H[E]$.[22] Hence the metric in $H[E]$ is topologically equivalent to the metric $\max_{x \in E} \rho_E(gx\, , \, hx)$ usually used to metrize the space of continuous mappings of E into itself, but with respect to the latter metric the space of automorphisms is not complete.

The main object of our investigation, however, will not be the space $H[E]$ but rather the subspace consisting of all measure-preserving automorphisms with respect to a given LS measure μ.[23] This subspace, with the same metric as in $H[E]$, will be denoted by $M[E, \mu]$. To see that $M[E, \mu]$ is a closed subset of $H[E]$, and therefore complete, consider any sequence of measure-preserving automorphisms T_n converging to the limit T in $H[E]$. Let F be any closed subset of E, and G any open set containing TF. From the uniform convergence of the sequence T_n it follows that $T_nF \subset G$, for all sufficiently large n, so that $\mu^*G \geqq \mu^*T_nF = \mu^*F$. Therefore $\mu^*F \leqq \mu^*TF$, by condition M5. Similar reasoning applied to the sequence T_n^{-1} yields the inverse inequality, so that $\mu^*TF = \mu^*F$ for all closed sets F. Hence equality holds also for all open sets, therefore for all sets, and the limiting automorphism T is therefore measure-preserving. Since the measure-preserving automorphisms of E form a group, it follows that $M[E, \mu]$ is a metric group, in fact, a closed subgroup of $H[E]$.

In case E is a rectangular r-cell R, and μ is the ordinary r-dimensional Lebesgue measure m in R, we shall write simply $M[R]$ instead of $M[R, m]$. The closed subgroups of $M[R]$ and $H[R]$ consisting of automorphisms that leave all boundary points fixed, we shall denote by $M_0[R]$ and $H_0[R]$ respectively.

THEOREM 1: *Let E be any regularly connected polyhedron of dimension $r \geqq 2$,*

---

[21] S. Banach, *Théorie des opérations linéaires*, Warsaw 1932, Chap. 1 and p. 229. Our metric in $H[E]$ is topologically equivalent to the one there assigned to this group.

[22] J. Schreier and S. Ulam, *Über topologische Abbildungen der euklidischen Sphären.* Fund. Math. **23** (1934) 102–118, esp. p. 104.

[23] We shall denote general measures by μ or ν. The letter m will be reserved for Lebesgue measure. General automorphisms will be denoted by g or h, measure-preserving automorphisms by S or T. For LS measures, the definition of a measure-preserving automorphism may be taken to be the condition $\mu^*TA = \mu^*A$ for every A.

and let μ be any r-dimensional Lebesgue-Stieltjes measure in E. In the space $M[E, \mu]$ of measure-preserving automorphisms of E the metrically transitive automorphisms form a residual G_δ set.[24]

The proof of this theorem will only be completed in Part III. In Part IV we shall discuss the generality of the result and its significance for metrically transitive flows.

In a previous note[25] one of us has obtained a similar result for the set of *topologically transitive* automorphisms, for slightly different assumptions on E. A transformation is topologically transitive if there do not exist two disjoint invariant open sets, both non-void, (or, equivalently, if the sequence of images of some point is dense in E). Metrical transitivity with respect to an r-dimensional LS measure obviously implies topological transitivity. Hence Theorem 1 may be considered as a generalization of the earlier result, but whereas the latter required only the most elementary constructions and properties of measure-preserving automorphisms, the present theorem will require a much more extensive investigation.

As regards the interpretation of Theorem 1 as a proof of the conjecture that metrical transitivity is the general case, it is interesting to note that the topological notion of probability in the sense of category can be subsumed under the general theory of measure and probability. In any complete metric space, if we define $\mu^*A = 0$ or 1 according as A is of first or second category, it is easily verified that conditions M1, M2, M3 are satisfied, and that the class of measurable sets consists of the first category sets and the residual sets, which have measures zero and one respectively. The notion of probability in the sense of category is therefore a special case of the general notion of (non-Borel) measure.

2. Preliminary Lemmas

The object of this first sequence of lemmas is to show that the general situation can be reduced, in all essential respects, to the consideration of ordinary Lebesgue measure in a rectangular r-cell. In making this reduction, a central role is played by the notion of a continuous map f of a convex cell Z onto a polyhedron E which is a *homeomorphism up to the boundary*. By this we shall mean that f maps the interior Z_0 of Z homeomorphically onto fZ_0. For such continuous maps $fBdZ = E - fZ_0$, so that $f^{-1}f$ is single-valued on Z_0, though possibly multiple-valued on BdZ.

LEMMA 1: *Let E be the polyhedron of a regularly connected euclidean complex K of dimension $r \geqq 1$. It is possible to represent E as the continuous image of a convex r-cell Z under a map f which is a homeomorphism up to the boundary and which is a simplicial map of a certain subdivision of Z onto K.*

This is not a new result, but rather a corollary of the known result that any

[24] A G_δ set is one that can be represented as a countable intersection of open sets.

[25] J. C. Oxtoby, *Note on transitive transformations*. Proc. Nat. Acad. USA. **23** (1937) 443–446. See also E. Hopf, *Statistische Probleme und Ergebnisse in der klassischen Mechanik*. Actualités Scientifiques et Industrielles, No. 737 (1938) 5–16.

regularly connected complex can be obtained from a convex cell by suitable boundary identifications.[26] To prove it, let $\sigma_1, \cdots, \sigma_n$ be the base simplexes of K, so numbered that each, after the first, has at least one regular face $\sigma_{i-1,i}$ in common with one of its predecessors. The possibility of such a numbering follows at once from the definition of a regularly connected complex.[27] Let τ_1 be any simplex in r-space, and set its vertices in correspondence with those of σ_1. Let $\tau_{1,2}$ be the face of τ_1 that corresponds to $\sigma_{1,2}$. Take a new vertex outside τ_1 which forms with $\tau_{1,2}$ a simplex τ_2 such that $\tau_1 + \tau_2$ is still convex. Any point sufficiently near the center of $\tau_{1,2}$ will suffice. Let this vertex correspond to the remaining vertex of σ_2. Evidently we have a simplicial map of the complex $\tau_1 + \tau_2$ onto the complex $\sigma_1 + \sigma_2$, and the only $(r - 1)$-simplex in $\tau_1 + \tau_2$ which does not lie on the boundary is $\tau_{1,2}$. Proceeding by induction, suppose that we have a complex K_i with r-dimensional base simplexes $\tau_1, \cdots, \tau_i$ whose union is a convex cell, and a simplicial map onto the complex $\sigma_1 + \sigma_2 + \cdots + \sigma_i$ under which τ_j corresponds to σ_j, $1 \leqq j \leqq i$. Also suppose that the only $(r$-$1)$-simplexes of K_i that do not lie on the boundary are $\tau_{1,2}, \cdots, \tau_{i-1,i}$ which correspond to $\sigma_{1,2}, \cdots, \sigma_{i-1,i}$. The simplex $\sigma_{i,i+1}$ is a face of some one of $\sigma_1, \cdots, \sigma_i$. Let $\tau_{i,i+1}$ be the corresponding face of the corresponding one of the simplexes $\tau_1, \cdots, \tau_i$. Since $\sigma_{i,i+1}$ is regular, it is distinct from $\sigma_{1,2}, \cdots, \sigma_{i-1,i}$, and it follows that $\tau_{i,i+1}$ lies on the boundary of the convex cell. We can therefore adjoin a new simplex τ_{i+1} having $\tau_{i,i+1}$ for a face, and such that $\tau_1 + \cdots + \tau_{i+1}$ is again convex. If we map the new vertex of τ_{i+1} onto the remaining vertex of σ_{i+1} the simplicial map will be extended to map $K_{i+1} = \tau_1 + \cdots + \tau_{i+1}$ onto $\sigma_1 + \cdots + \sigma_{i+1}$, and the only $(r - 1)$-dimensional simplexes of K_{i+1} not on its boundary will be $\tau_{1,2}, \cdots, \tau_{i,i+1}$. At the n-th stage we obtain a simplicial map of a complex K_n onto K. If we map the simplexes $\tau_1, \cdots, \tau_n$ affinely onto $\sigma_1, \cdots, \sigma_n$, we obtain a continuous map of a convex cell Z onto E. The interior of Z is the union of the interiors of $\tau_1, \cdots, \tau_n$ and of $\tau_{1,2}, \cdots, \tau_{n-1,n}$. These correspond to the interiors of $\sigma_1, \cdots, \sigma_n$ and of $\sigma_{1,2}, \cdots, \sigma_{n-1,n}$ respectively, which are disjoint since these simplexes are all distinct. Hence distinct interior points of Z go into distinct points, and the map is therefore a homeomorphism up to the boundary.

LEMMA 2: *Let f be a continuous map of a rectangular r-cell R, $r \geqq 1$, onto a polyhedron E which is a homeomorphism up to the boundary. Let μ be an r-dimensional Lebesgue-Stieltjes measure in E, and suppose that $\mu fBdR = 0$. Then the function $\nu^*A = \mu^*fA$ defines an r-dimensional Lebesgue-Stieltjes measure in R.*

That ν^* satisfies conditions M1 to M3 may be verified by inspection. Con-

[26] See AH p. 264.

[27] A complex is *homogeneous r-dimensional* if every simplex lies on an r-simplex. A *regular face* is an $(r - 1)$-simplex that lies on two and only two r-simplexes of the complex. A homogeneous r-dimensional complex K is *regularly connected* if in any division of K into two r-dimensional subcomplexes, these have in common at least one regular face. It can be shown that a polyhedron is regularly connected if and only if it is the polyhedron of a regularly connected complex. See AH p. 402.

dition M4 is less evident in view of the fact that f need not be 1:1. To verify it, consider any two sets A, B contained in R and separated by a positive distance. Their closures $\bar{A}$, $\bar{B}$ are disjoint, and therefore $f\bar{A}$ and $f\bar{B}$ intersect at most in a subset of $fBdR$, since all other points have unique antecedents. Hence $\mu f\bar{A} \cdot f\bar{B} = 0$, by hypothesis. It is possible to enclose $f(A + B)$ in a G_δ set C such that $\mu^*f(A + B) = \mu C$. Let $A_1 = C \cdot f\bar{A}$ and $B_1 = C \cdot f\bar{B}$. Then we have $fA \subset A_1$ and $fB \subset B_1$, and so $\mu^*f(A + B) = \mu^*(fA + fB) \leqq \mu^*(A_1 + B_1) \leqq \mu^*C = \mu^*f(A + B)$, that is, $\mu^*f(A + B) = \mu(A_1 + B_1)$. But A_1 and B_1 are Borel sets that intersect in a set of measure zero, therefore $\mu(A_1 + B_1) = \mu A_1 + \mu B_1 \geqq \mu^*fA + \mu^*fB$, and so $\nu^*(A + B) \geqq \nu^*A + \nu^*B$. The inverse inequality follows from M2, and so M4 is satisfied. To verify condition M5, we have $\nu^*A = \mu^*fA = \inf_{G \supset fA} \mu G$. But whenever G is an open set containing fA, $f^{-1}G$ is an open set in R containing A. Hence $\nu^*A \geqq \inf_{G \supset A} \nu G$. Again the inverse inequality follows from M2, and the verification that ν is a LS measure is complete. Furthermore, ν is positive for non-void open sets in R, because the image of such a set contains a neighborhood of a regular point of E. Finally, it is evident that single points have measure zero, and by hypothesis $\nu BdR = \mu fBdR = 0$. Hence ν is an r-dimensional LS measure in R.

LEMMA 3: *Let μ be an r-dimensional LS measure in a regularly connected polyhedron E, $r \geqq 1$. It is possible to represent E as the continuous image of a rectangular r-cell R under a map f which is a homeomorphism up to the boundary, and which is such that $\mu^*fA = m^*A$ for all $A \subset R$, where m^* denotes ordinary Lebesgue outer measure in R.*

E is homeomorphic to the polyhedron E_1 of a euclidean complex K_1. It is possible to choose the correspondence in such a way that all $(r - 1)$-dimensional simplexes of K_1 correspond to sets of measure zero. Because if μ_1 is the measure in E_1 corresponding to μ under any homeomorphism, the singular $(r - 1)$-simplexes have measure zero by hypothesis, and a suitably chosen automorphism of E_1 will displace all regular $(r - 1)$-simplexes into sets of μ_1-measure zero. (First displace the interiors of all 1-simplexes not contained in the set of singular points to nearby positions having measure zero. Then do the same for 2-simplexes, and so on until all regular $(r - 1)$-simplexes have been displaced to sets having μ_1-measure zero. The automorphism need displace only regular points of E_1 and may leave all vertices of K_1 fixed.[28]) We may therefore suppose that E is the polyhedron of a euclidean complex K and that all lower dimensional simplexes of K have μ-measure zero. By Lemma 1 there exists a continuous map of a convex cell Z onto E which is a homeomorphism up to the boundary and is a simplicial map of a certain subdivision of Z onto K. Let us combine this map with a homeomorphism of Z onto a rectangular r-cell R, of volume $mR = \mu E$,

[28] More generally, any set of first category can be displaced to a set of measure zero by an automorphism. See our joint paper, *On the equivalence of any set of first category to a set of measure zero.* Fund. Math. **31** (1938) 201–206.

and let f denote the resulting map of R onto E. Then f is a homeomorphism up to the boundary, and since BdR corresponds to certain $(r-1)$-dimensional simplexes of K, we see that $\mu f BdR = 0$. By Lemma 2, the function $\nu^*A = \mu^*fA$ defines an r-dimensional LS measure in R, and also we have $\nu R = mR$. In Part II (Theorem 2) it will be shown that for any such measure there exists an automorphism h of R such that $\nu^*A = m^*hA$ for all $A \subset R$. Assuming this result for the moment, we conclude that fh^{-1} is a continuous map of R onto E which is a homeomorphism up to the boundary, and that $m^*A = \mu^*fh^{-1}A$ for all $A \subset R$, as required.

LEMMA 4: *Let f be a map of R onto E with the properties stated in Lemma 3. Any Lebesgue-measure-preserving automorphism T of R that leaves the boundary fixed corresponds to a μ-measure-preserving automorphism fTf^{-1} of E. On the other hand, any μ-measure-preserving automorphism T of E defines a transformation $f^{-1}Tf$ in R which is a Lebesgue-measure-preserving homeomorphism of the open set $R_0 \cdot f^{-1}T^{-1}fR_0$ onto the open set $R_0 \cdot f^{-1}TfR_0$, both of which are contained in the interior R_0 of R and have measure equal to mR.*

To prove the first assertion we first show that fTf^{-1} is a 1:1 map of E onto itself. Consider any point p in fR_0. Then $f^{-1}p$ is single-valued, and therefore also $fTf^{-1}p$. If $p \notin fR_0$, then $p \epsilon fBdR$, and so $fTf^{-1}p = p$. Any point of $fBdR$ is its own image, and any point $p \epsilon fR_0$ is the image of $fT^{-1}f^{-1}p$. Hence fTf^{-1} is a 1:1 map of E onto itself, with inverse $fT^{-1}f^{-1}$. Both these transformations carry closed sets into closed sets, and so they are automorphisms. That fTf^{-1} preserves μ-measure is immediate, since $\mu^*A = m^*f^{-1}A = m^*Tf^{-1}A = \mu^*fTf^{-1}A$ for every set $A \subset E$.

To prove the second assertion of the lemma, consider any $p \in R_0 \cdot f^{-1}T^{-1}fR_0$. Then $Tfp \in fR_0$ and so $f^{-1}Tfp$ is single-valued on this domain. Its inverse is $f^{-1}T^{-1}f$ with domain $R_0 \cdot f^{-1}TfR_0$. Both of these are continuous throughout these domains. Thus $f^{-1}Tf$ maps the open set $R_0 \cdot f^{-1}T^{-1}fR_0$ homeomorphically onto $R_0 \cdot f^{-1}TfR_0$, and does so in a measure-preserving manner, as follows from the measure-preserving properties of f and T.

The partial correspondence between measure-preserving automorphisms of E and R described in Lemma 4 makes it possible to reduce the proof of Theorem 1 to the following lemma concerning transformations in R and ordinary Lebesgue measure.

LEMMA 5: *Let R_0 be the interior of a rectangular r-cell R, $r \geqq 2$, and let T be a measure-preserving homeomorphism of an open set $G \subset R_0$ onto an open set $TG \subset R_0$, where $mG = mR$, and let $\sigma_1, \cdots, \sigma_N$ be the cells of any given dyadic subdivision. There exist arbitrarily small automorphisms h_1 and h_2 of R that leave the boundary fixed such that h_1Th_2 is a measure-preserving homeomorphism of $h_2^{-1}G$, which has measure mR, onto h_1TG, and such that under this transformation a certain closed set F is transformed in the following manner: There exists a positive integer K such that the first KN images of F under iteration of h_1Th_2 are disjoint and exactly K are contained in the interior of each cell σ_i and contain exactly half the measure of σ_i.*

We shall prove this lemma in §7. We proceed to show that from it we can deduce

Theorem 1, as regards category. Let f be the map of R onto E defined in Lemma 3, and let $\sigma_1, \sigma_2, \cdots$, be an enumeration of all dyadic cells in R. If σ_i and σ_j belong to the same dyadic subdivision, let $E_{i,j}$ be the set of all automorphisms $T \in M[E, \mu]$ such that for some Borel set A we have $TA = A$, $\mu(A \cdot f\sigma_i) > \frac{3}{4}\mu f\sigma_i$, and $\mu(A \cdot f\sigma_j) < \frac{1}{4}\mu f\sigma_j$, otherwise undefined. Every metrically intransitive automorphism in $M[E, \mu]$ belongs to one of the sets $E_{i,j}$. Because if T is metrically intransitive, there exists a measurable set, and therefore a Borel set A, such that $TA = A$ and $0 < \mu A < \mu E$. Therefore $0 < mf^{-1}A < mR$. Let p, q be points of R at which the Borel set $f^{-1}A$ has metric density 1 and 0 respectively. It follows from Lebesgue's density theorem[29] that in any sufficiently fine dyadic subdivision the cells σ_i and σ_j that contain p and q respectively are such that $m(f^{-1}A \cdot \sigma_i) > \frac{3}{4}m\sigma_i$ and $m(f^{-1}A \cdot \sigma_j) < \frac{1}{4}m\sigma_j$. Therefore $\mu(A \cdot f\sigma_i) > \frac{3}{4}\mu f\sigma_i$ and $\mu(A \cdot f\sigma_j) < \frac{1}{4}\mu f\sigma_j$, so that T belongs to $E_{i,j}$. Thus, to prove that the metrically transitive automorphisms form a residual set in $M[E, \mu]$ it suffices to show that each of the sets $E_{i,j}$ is nowhere dense, because there are only countably many of them and their union contains all metrically intransitive automorphisms in $M[E, \mu]$.

Consider any $T \in M[E, \mu]$ and any set $E_{i,j}$. We shall show that arbitrarily near T there exists an automorphism S which together with a neighborhood lies outside $E_{i,j}$. According to Lemma 4, the transformation $f^{-1}Tf$ is a measure-preserving homeomorphism of the open set $G = R_0 \cdot f^{-1}T^{-1}fR_0$ onto another open set contained in R_0 and likewise having measure mR. Applying Lemma 5 to this transformation, there exist arbitrarily small automorphisms h_1, h_2 of R which leave the boundary fixed, such that $h_1f^{-1}Tfh_2$ is a measure-preserving homeomorphism of $h_2^{-1}G$ onto another open set with measure mR. Also there is a closed set F whose first KN images under this transformation are disjoint and equally distributed among the cells $\tau_1, \cdots, \tau_N$ of the dyadic subdivision to which σ_i and σ_j belong, and contain half their measure. By Lemma 4, the transformations fh_1f^{-1} and fh_2f^{-1} are automorphisms of E. Hence $S = (fh_1f^{-1})T(fh_2f^{-1})$ is an automorphism of E, and from the properties of $h_1f^{-1}Tfh_2$ it follows that S transforms the closed set fF in such a way that its first KN images are disjoint and equally distributed among $f\tau_1, \cdots, f\tau_N$, and contain half their measure. Now consider any Borel set A invariant under S. Let α be the fraction of the measure of fF contained in A. Suppose $\alpha \leqq \frac{1}{2}$, then the K disjoint images of fF in σ_i contain the same fraction of the measure of A, and therefore $mA\sigma_i \leqq \left(\frac{1}{2} + \frac{\alpha}{2}\right)m\sigma_i \leqq \frac{3}{4}m\sigma_i$, so that S does not belong to $E_{i,j}$. On the other hand, if $\alpha \geqq \frac{1}{2}$ the K disjoint images of fF in σ_j contain measure $\frac{\alpha}{2}m\sigma_j$, and therefore $mA\sigma_j \geqq \frac{\alpha}{2}m\sigma_j \geqq \frac{1}{4}m\sigma_j$, so that in this case also S does not belong to $E_{i,j}$. Furthermore, the first KN images of fF under any automorphism sufficiently close to S will be distributed among $f\tau_1, \cdots, f\tau_N$ in the same

[29] See e.g. Hobson, *The theory of functions of a real variable*, 2nd edition, Cambridge 1921, vol. 1, p. 181.

way as under S, so that the same reasoning shows that a whole neighborhood of S lies outside $E_{i,j}$. But h_1 and h_2 were arbitrarily small, hence also the automorphisms fh_1f^{-1} and fh_2f^{-1}, so that S is arbitrarily close to T. Hence $E_{i,j}$ is nowhere dense in $M[E, \mu]$.

It will be observed that S may be written in the form S_1T, where S_1 is the measure-preserving transformation ST^{-1}. But S_1 may also be written $(fh_1f^{-1})T(fh_2f^{-1})T^{-1}$, from which it follows that S_1 leaves all singular points fixed. We have therefore proved the more precise result that $E_{i,j}$ is nowhere dense with respect to every coset of the closed normal subgroup of automorphisms that leave all singular points fixed. Since these cosets are complete spaces, we may state Theorem 1 in the following slightly stronger form which we shall need later.

COROLLARY: *For every $T \in M[E, \mu]$ there exists an arbitrarily small automorphism $S \in M[E, \mu]$, that leaves all singular points fixed, such that ST is metrically transitive with respect to μ.*

II. MEASURES TOPOLOGICALLY EQUIVALENT TO LEBESGUE MEASURE

3. The Fundamental Theorem

Let μ be any measure function in a space, and h an automorphism of the space. The function μhA, considered as defined for all sets such that hA is measurable, is easily seen to be a measure function. It would naturally be described as a measure automorphic to μ, (or equivalent to μ under automorphism). It is easily seen that any measure automorphic to a LS measure is again a LS measure, and that the same transformation also effects a correspondence between their outer measures. Likewise, any measure automorphic to an r-dimensional measure is r-dimensional. The object of the present Part will be to obtain a simple characterization of measures automorphic to Lebesgue measure. The basic result, stated as Theorem 2, was originally proposed by one of us in 1936, in connection with some other group-theoretic investigations, and a proof was obtained at that time by J. von Neumann, but was not published. The present proof, based on somewhat different considerations, was worked out subsequently. The result is here published for the first time.

THEOREM 2: *In order that a measure μ in a rectangular r-cell R, $r \geqq 1$, be automorphic to Lebesgue measure it is necessary and sufficient that it be an r-dimensional Lebesgue-Stieltjes measure and that $\mu R = mR$. The correspondence can always be effected by an automorphism that leaves the boundary of R fixed.*

An equivalent formulation, from the standpoint of the Carathéodory theory, runs as follows.

THEOREM 2_1: *In order that a function μ^* defined for all subsets of R be automorphic to the Lebesgue outer measure m^* it is necessary and sufficient that it satisfy conditions $M1$ to $M5$, and also*

M6: $\mu^*G > 0$ *if G is non-void, open*

M7: $\mu^*p = 0$ *for every point p*

M8: $\mu^*BdR = 0$

and finally that $\mu^*R = mR$. *If* μ^* *satisfies these conditions there exists an automorphism* h *of* R *such that* $\mu^*A = m^*hA$ *for every* $A \subset R$, *and such that* h *leaves the boundary fixed.*

The equivalence of the two formulations is evident. Before proving the theorem we shall derive from it a number of corollaries and generalizations. In Part V, §13 to §15 and §17, we shall give some further applications.

COROLLARY 1: *Let* E *be any regularly connected polyhedron of dimension* $r \geqq 1$, *and let* μ_1 *and* μ_2 *be two r-dimensional LS measures in* E *such that* $\mu_1 E = \mu_2 E$. *There exists an automorphism* h *such that* $\mu_1^* A = \mu_2^* hA$ *for all* $A \subset E$, *and such that* h *leaves all singular points of* E *fixed. In particular, any r-dimensional LS measure in a regularly connected euclidean polyhedron* E *is automorphic to the Lebesgue measure in* E, *provided only that* $\mu E = mE$.

Let ν be the r-dimensional LS measure in E defined by $\nu^*A = \frac{1}{2}(\mu_1^*A + \mu_2^*A)$, and let f be the map of a rectangular r-cell R onto E given by Lemma 3. Then $m^*A = \nu^*fA$, and in particular, $\nu fBdR = 0$, so that we have $\mu_1 fBdR = \mu_2 fBdR = 0$. Let ν_1^* and ν_2^* be the outer measures in R defined by $\nu_1^*A = \mu_1^*fA$ and $\nu_2^*A = \mu_2^*fA$. By Lemma 2, these are both r-dimensional LS measures with $\nu_1 R = \nu_2 R = mR$. Hence, by Theorem 2_1, there exist automorphisms h_1, h_2 of R that leave the boundary fixed, such that $\nu_1^*A = m^*h_1A$ and $\nu_2^*A = m^*h_2A$. Then $h = fh_2^{-1}h_1f^{-1}$ is an automorphism of E that leaves singular points fixed and carries μ_1^* into μ_2^* as required.

COROLLARY 2: *Let* E *be a regularly connected polyhedron of dimension* $r \geqq 1$, *and* E_1 *any homeomorphic euclidean polyhedron. The r-dimensional Lebesgue-Stieltjes measures in* E *are the same as the measures of the form* $C \cdot mhA$, *where* h *is an arbitrary homeomorphism of* E *onto* E_1, C *is an arbitrary positive constant, and* m *denotes Lebesgue measure in* E.

This follows at once from Corollary 1 and the fact that any multiple of a measure homeomorphic to an r-dimensional measure is again an r-dimensional measure.

COROLLARY 3: *Let* E_1 *and* E_2 *be homeomorphic regularly connected polyhedra of dimension* $r \geqq 1$, *and let* μ_1 *and* μ_2 *be r-dimensional Lebesgue-Stieltjes measures in* E_1 *and* E_2 *respectively, such that* $\mu_1 E = \mu_2 E$. *Let* h *be any homeomorphism of* E_1 *onto* E_2. *There exists a homeomorphism* g *of* E_1 *onto* E_2 *which carries* μ_1 *into* μ_2 *and which is equal to* h *for all singular points. In particular, there exists a measure-preserving automorphism of a rectangular r-cell which is equal to any given automorphism on the boundary.*

The measure $\nu_1 A = \mu_2 hA$ is an r-dimensional LS measure in E_1, and $\nu_1 E_1 = \mu_1 E_1$. By Corollary 1, there exists an automorphism h_1 of E_1 such that $\mu_1 A = \nu_1 h_1 A$, and such that it leaves singular points fixed. Hence $g = hh_1$ is a homeomorphism of E_1 onto E_2 that takes μ_1 into μ_2 and is equal to h for all singular points.

For later use the following corollary is convenient.

COROLLARY 4: *Let* μ *be a Lebesgue-Stieltjes measure in a rectangular r-cell* R, $r \geqq 2$, $\mu R = mR$, *and let* L *be any straight line segment contained in the interior*

of R such that $\mu L = 0$. Then the automorphism that carries μ into m may be so chosen that it leaves L fixed, as well as the boundary.

To see this, let R_1 and R_2 be rectangular r-cells with R_1 contained in the interior of R_2, and let E be the polyhedron obtained from R_2 by removing the interior of R_1. In the case $r \geqq 2$ it is evident that E is regularly connected. From E we can obtain R by identifying points of the inner boundary that have the same x_1 coordinate, say. That is, we can define a continuous map f of E onto R which carries the inner boundary into L and elsewhere is 1:1. The two measures $\mu_1 A = mfA$ and $\mu_2 A = \mu f A$ are easily verified to be r-dimensional LS measures in E, and $\mu_1 E = \mu_2 E$. By Corollary 1, there exists an automorphism h which leaves the inner and outer boundary of E fixed and carries μ_1 into μ_2. Hence fhf^{-1} is an automorphism of R that carries μ into m and leaves both L and BdR fixed.

The next two corollaries characterize measures automorphic to Lebesgue measure in the whole of euclidean space $E^{(r)}$.

COROLLARY 5: *A function μ^* defined for all subsets of $E^{(r)}$, $r \geqq 2$, is automorphic to Lebesgue outer measure m^* if and only if it satisfies the conditions stated in Theorem 2_1, except that conditions* M1 *and* M8 *are to be replaced by*

M1′: $0 \leqq \mu^* A \leqq \mu^* E^{(r)} = +\infty$

M8′: *$\mu^* A < +\infty$ for every bounded set A.*

The necessity of each of the eight conditions is evident. To prove that they are sufficient, let μ be the measure defined by any such outer measure. Let R_α be the closed cube in $E^{(r)}$ with edges parallel to the axes, center at the origin, and edge length α. Under our hypotheses, the function $f(\alpha) = \mu R_\alpha$ is a finite, strictly increasing function in $0 < \alpha < +\infty$. Such a function can have at most countably many discontinuities, hence we can select a sequence of values $0 < \alpha_1 < \alpha_2 < \cdots$, tending to infinity, at each of which $f(\alpha)$ is continuous. This means that $\mu BdR_{\alpha_i} = 0$. By hypothesis we have also that $\mu R_{\alpha_i} \to +\infty$ as $i \to \infty$. Let β_i be the edge length such that $mR_{\beta_i} = \mu R_{\alpha_i}$. Then $0 < \beta_1 < \beta_2 < \cdots$, and $\beta_i \to +\infty$ as $i \to \infty$. We can therefore define a radial transformation h_1 of $E^{(r)}$ to carry R_{α_i} into R_{β_i}, $i = 1, 2, \cdots$. Consider the new outer measure $\mu_1^* A = \mu^* h_1^{-1} A$. Then $\mu_1 R_{\beta_i} = mR_{\beta_i}$ and $\mu_1 BdR_{\beta_i} = 0$, $i = 1, 2, \cdots$. Let $E_1 = R_{\beta_1}$ and for $n > 1$ let E_n be R_{β_n} minus the interior of $R_{\beta_{n-1}}$. Then each E_n is a regularly connected polyhedron, since we have assumed $r \geqq 2$, and μ_1 is an r-dimensional measure in it such that $\mu_1 E_n = mE_n$. By Corollary 1, there exists an automorphism h_2 of E_n such that $\mu_1^* A = m^* h_2 A$, for all $A \subset E_n$. Since h_2 leaves the inner and outer boundaries of E_n fixed, these automorphisms join up to form an automorphism of $E^{(r)}$. In view of the measurability of the sets E_n and the fact that their boundaries have measure zero, we have, for any set $A \subset E^{(r)}$,

$$\mu_1^* A = \sum_{n=1}^{\infty} \mu_1^* A E_n = \sum_{n=1}^{\infty} m^* h_2(AE_n) = \sum_{n=1}^{\infty} m^*(h_2 A) E_n = m^* h_2 A.$$

Therefore $\mu^* A = m^* h_2 h_1 A$ for all $A \subset E^{(r)}$.

It should be added that in the case $r = 1$ the conditions stated in Corollary 5 are not sufficient, but a complete characterization is obtained by merely adding the requirement that the positive and negative halves of the line both have infinite measure. The function $f(x) = \mu(0 \leqq t \leqq x)$ for $x \geqq 0$, $f(x) = -\mu(x \leqq t \leqq 0)$ for $x < 0$ then defines an automorphism of the line that carries μ into m.

COROLLARY 6: *A measure in $E^{(r)}$, $r \geqq 2$, is automorphic to Lebesgue measure if and only if it is an infinite Lebesgue-Stieltjes measure that is zero for points and positive for non-void open sets.*

Here we may understand by a LS measure one that is derived from a non-negative additive function of an interval. Such a measure is necessarily finite for bounded sets, and so condition M8′ is implicit. The rest then follows from Corollary 5.

4. Proof of Theorem 2

It may be remarked first of all that in the case $r = 1$ Theorem 2 is almost trivial. For suppose R is the unit interval $0 \leqq x \leqq 1$, and let $hx = \mu(0 \leqq t \leqq x)$. Then hx is strictly increasing, in virtue of the additivity of μ and the fact that all subintervals have positive measure. It is also continuous, since points have measure zero. As x describes the interval $(0, 1)$, hx describes the same interval. Hence h is an automorphism of R. By definition, we have $\mu A = mhA$ for every interval A of the form $(0, x)$. From additivity it follows that equality holds for all intervals, and therefore for all open sets. But this implies that $\mu^*A = m^*hA$ for all sets A, by condition M5. Hence h effects the desired transformation, and also leaves the end points fixed.

However, even in the case $r = 2$ the difficulties of the general case already appear. Our proof will be based on a sequence of lemmas whose motivation lies in the idea of securing first that $\mu A = mhA$ for all sets of a division automorphic to a dyadic subdivision. Then h is modified within each of these sets so as to secure equality for the sets of a finer subdivision. Finally, a convergent sequence of such modifications is obtained and the limiting automorphism effects the desired transformation for all sets.

LEMMA 6: *Let μ be any Lebesgue-Stieltjes measure in R that is zero for points and for the boundary, and let α be any number in the interval $0 < \alpha < \mu R$. There exists an open set G contained in the interior of R such that $\mu G = \alpha$.*

Since μR is finite, there can be at most countably many planes parallel to the faces of R that intersect R in sets of positive μ-measure. Hence we can divide R into a finite number of rectangular r-cells $\sigma_1, \cdots, \sigma_N$ of diameter less than $\frac{1}{2}$, whose boundaries all have μ-measure zero. Let i be the least integer such that $\mu(\sigma_1 + \sigma_2 + \cdots + \sigma_i) \geqq \alpha$, and let G_1 be the union of the interiors of $\sigma_1, \sigma_2, \cdots, \sigma_{i-1}$. (Take G_1 equal to the void set in case $i = 1$). Then $\mu G_1 < \alpha$, but $\mu G_1 + \mu\sigma_i \geqq \alpha$. Now consider the cell σ_i, calling it R_1, and divide it into rectangular r-cells $\sigma_1^{(1)}, \sigma_2^{(1)}, \cdots, \sigma_{N_1}^{(1)}$ of diameter less than $\frac{1}{4}$, whose boundaries all have μ-measure zero. Again we find an open set $G_2 \subset R_1$, either void or

consisting of the interiors of some of the cells $\sigma_i^{(1)}$, such that $\mu(G_1 + G_2) < \alpha$, while $\mu(G_1 + G_2) + \mu R_2 \geqq \alpha$, where R_2 is one of the cells $\sigma_i^{(1)}$ and is disjoint to $G_1 + G_2$. Proceeding in this manner, we find disjoint open sets $G_1, G_2, \cdots$ interior to R, and a nested sequence of rectangular r-cells $R_1, R_2, \cdots$ whose diameters tend to zero, such that $\alpha - \mu R_n \leqq \mu(G_1 + G_2 + \cdots + G_n) < \alpha$, $n \geqq 1$. The cells R_n intersect in a point p, and we have $\lim_{n\to\infty} \mu R_n = \mu p = 0$, by hypothesis. Hence $\mu G = \alpha$, where G is the union of the sets G_n.

LEMMA 7: *Let μ be any Lebesgue-Stieltjes measure in R that is zero for points and for the boundary. Let R_1 and R_2 be the two cells obtained by bisecting R perpendicularly to one of its edges. Let α_1 and α_2 be any two positive numbers such that $\alpha_1 + \alpha_2 = \mu R$. There exists an automorphism $h \in H_0[R]$ such that $\mu h R_1 = \alpha_1$ and $\mu h R_2 = \alpha_2$.*

We shall establish the existence of h by a category argument. Let H_1 be the set of automorphisms h in $H_0[R]$ such that $\mu h R_1 \geqq \alpha_1$ and $\mu h R_2 \geqq \alpha_2$. This is a closed set, for if h_n is any sequence of automorphisms in H_1 tending uniformly to h, then any ϵ-neighborhood of hR_i, $i = 1, 2$, contains $h_n R_i$ for all sufficiently large n, and therefore has μ-measure at least α_i. This being true for every $\epsilon > 0$, it follows that $\mu h R_1 \geqq \alpha_1$ and $\mu h R_2 \geqq \alpha_2$. Furthermore, the set H_1 is non-void. For, unless it contains the identity, we have either $\mu R_1 < \alpha_1$ or $\mu R_2 < \alpha_2$. Suppose $\mu R_1 < \alpha_1$. We can deform R by a continuous family of automorphisms $h_\lambda \in H_0[R]$ in such a way that as λ increases, $h_\lambda R_1$ includes more and more of the interior of R_2. At the limiting value λ_0 where $\mu h_\lambda R_1$ first becomes greater than or equal to α_1 we have also $\mu h_\lambda R_2 \geqq \alpha_2$, because $\mu h_{\lambda_0} R_2 = \lim_{\lambda\to\lambda_0^-} \mu h_\lambda R_2$, and for $\lambda < \lambda_0$ we have $\mu h_\lambda R_2 \geqq \mu R - \mu h_\lambda R_1 > \mu R - \alpha_1 = \alpha_2$. Hence the set H_1, as a non-void closed subset of a complete space, is itself a complete space. We shall show that in it the set of automorphisms that fulfill the requirements of the lemma is residual.

Let E_n be the set of all h in H_1 such that $\mu h R_1 \geqq \alpha_1 + \frac{1}{n}$. This set is closed in H_1, since we have already seen that it is closed in $H_0[R]$. To show that it is nowhere dense in H_1 it suffices to show that if $h \in E_n$ there exists an arbitrarily small automorphism $g \in H_0[R]$ such that $hg \in H_1 - E_n$. Suppose that $h \in E_n$ and consider the new measure $\nu A = \mu h A$. Then $\nu R_1 \geqq \alpha_1 + \frac{1}{n}$ and $\nu R_2 \geqq \alpha_2$. Let R_3 denote the $(r - 1)$-dimensional face common to R_1 and R_2. Then $\nu R_1 + \nu R_2 - \nu R_3 = \nu R = \alpha_1 + \alpha_2$. Hence $\nu R_1 - \alpha_1 \leqq \nu R_1 - \alpha_1 + \nu R_2 - \alpha_2 = \nu R_3$, and this with the inequality $\nu R_1 \geqq \alpha_1 + \frac{1}{n}$ gives $0 < \nu R_1 - \alpha_1 - \frac{1}{2n} < \nu R_3$. Considered with respect to R_3, ν is a LS measure that is zero for points and for the boundary. Hence, by Lemma 6, there exists a set G, open with respect to R_3 and contained in its interior, such that $\nu G = \nu R_1 - \alpha_1 - \frac{1}{2n}$. Now let g be an automorphism of R that leaves the boundary fixed and also the points

of $R_3 - G$, but let it displace all points of G slightly into the interior of R_1. Such an automorphism can be found arbitrarily close to the identity. Since $gR_2 \supset R_2$, we have $\nu g R_2 \geqq \alpha_2$; and $\nu g R_1$ differs only slightly from $\nu R_1 - \nu G$, which is equal to $\alpha_1 + \frac{1}{2n}$. If g is sufficiently near the identity we shall have $\alpha_1 < \nu g R_1 < \alpha_1 + \frac{1}{n}$. Going back to the measure μ, we have $\alpha_1 < \mu h g R_1 < \alpha_1 + \frac{1}{n}$ and $\mu h g R_2 \geqq \alpha_2$. Hence $hg \in H_1 - E_n$, and so E_n is nowhere dense in H_1. Therefore the set $H_1 - \sum_1^\infty E_n$ is residual, that is, the equation $\mu h R_1 = \alpha_1$ holds for a residual set of automorphisms in H_1. The rôle of R_2 being similar, we see that the automorphisms h such that $\mu h R_2 = \alpha_2$ is also residual in H_1. The intersection of these two residual sets is the set of automorphisms having the properties required by the lemma, and therefore such automorphisms exist.

It may be remarked that in the proof of this lemma it is possible to confine attention to automorphisms that involve only a single coordinate, namely that in the direction along which R is bisected. Thus, if desired, the automorphism h can be chosen so as to leave all coordinates unchanged except this one.

LEMMA 8: *Let μ be any Lebesgue-Stieltjes measure in R that is zero for points and for the boundary. Let $\sigma_1, \cdots, \sigma_N$ be the cells of any dyadic subdivision of R, and let $\alpha_1, \cdots, \alpha_N$ be associated positive numbers whose sum is equal to μR. There exists an automorphism $h \in H_0[R]$ such that $\mu h \sigma_i = \alpha_i$, $i = 1, \cdots, N$.*

The proof consists in applying Lemma 7 a finite number of times. The cells of any dyadic subdivision are obtained from R by making a finite number of bisections. Let R_1, R_2 be the cells obtained from the first bisection. Let R_{11}, R_{12}; R_{21}, R_{22} be the cells obtained by bisecting R_1 and R_2. After n bisections we get 2^n cells $R_{i_1 \cdots i_n}$, $i_1, \cdots, i_n = 1, 2$, and for $n = K \equiv \log_2 N$ the cells $R_{i_1 \cdots i_K}$ are the cells $\sigma_1, \cdots, \sigma_N$ renamed. Let $\beta_{i_1 \cdots i_K}$ be the number α_i associated with $R_{i_1 \cdots i_K}$. For $1 \leqq n < K$ let $\beta_{i_1 \cdots i_n} = \sum \beta_{i_1 \cdots i_K}$ summed over $i_{n+1}, \cdots, i_K = 1, 2$. For $n = 1$ we have $\beta_1 + \beta_2 = \mu R$, and by Lemma 7 we can find h_1 such that $\mu h_1 R_i = \beta_i$, $i = 1, 2$. Suppose we have defined h_n such that $\mu h_n R_{i_1 \cdots i_n} = \beta_{i_1 \cdots i_n}$, $i_1, \cdots, i_n = 1, 2$. Form the new measure $\mu_n A = \mu h_n A$. Then μ_n satisfies the conditions of Lemma 7 with respect to each of the cells $R_{i_1 \cdots i_n}$, and since $\beta_{i_1 \cdots i_n 1} + \beta_{i_1 \cdots i_n 2} = \mu_n R_{i_1 \cdots i_n}$, we can find an automorphism $g_{i_1 \cdots i_n}$ in $H_0[R_{i_1 \cdots i_n}]$ such that $\mu_n g_{i_1 \cdots i_n} R_{i_1 \cdots i_{n+1}} = \beta_{i_1 \cdots i_{n+1}}$, $i_{n+1} = 1, 2$. Since $g_{i_1 \cdots i_n}$ leaves the boundary of $R_{i_1 \cdots i_n}$ fixed, these automorphisms join up to form an automorphism $g \in H_0[R]$, and we have $\mu h_n g R_{i_1 \cdots i_{n+1}} = \beta_{i_1 \cdots i_{n+1}}$. Setting $h_{n+1} = h_n g$, the induction hypotheses are again satisfied, and for $n = K$ we get an automorphism $h = h_K$ that fulfills the requirements of the lemma.

LEMMA 9: *Let μ be any Lebesgue-Stieltjes measure in R that is zero for points and for the boundary. There exists an automorphism $h \in H_0[R]$ such that for every dyadic cell σ we have $\mu h Bd\sigma = 0$.*

We shall show that the automorphisms having the desired property form a

residual set in $H_0[R]$.[30] Let R_1 be the $(r - 1)$-dimensional cell in which R is intersected by any one of the planes used in forming dyadic subdivisions. Let E_n be the set of automorphisms $h \epsilon H_0[R]$ such that $\mu h R_1 \geqq \frac{1}{n}$. This is a closed set. If h is any element of E_n, let $\nu A = \mu h A$. Choose $\epsilon > 0$ so that the ϵ-neighborhood of R_1 has ν-measure less than $\nu R_1 + \frac{1}{n}$. Let $g \epsilon H_0[R]$ be an automorphism that displaces the interior of R_1 into a set disjoint to R_1 but contained in its ϵ-neighborhood. Then, since the intersection of R_1 and gR_1 has ν-measure zero, we have $\nu R_1 + \nu g R_1 = \nu(R_1 + gR_1) \leqq \nu(\epsilon\text{-neighborhood of } R_1) < \nu R_1 + \frac{1}{n}$. Hence $\mu h g R_1 < \frac{1}{n}$, and $hg \epsilon H_0 - E_n$. Since hg is arbitrarily near h, the set E_n is nowhere dense, and $H_0 - \sum_1^\infty E_n$ is residual. The intersection of the residual sets corresponding to each of the countably many planes used in forming dyadic subdivisions is therefore also residual, and the automorphisms belonging to this set have the required property.

In the following lemma we require for the first time that the measures under consideration be positive for non-void open sets.

LEMMA 10: *Let μ, ν be two r-dimensional Lebesgue-Stieltjes measures in R such that $\mu R = \nu R$, and let $\epsilon > 0$ be given. There exist automorphisms g, h in $H_0[R]$ such that for each cell σ of a certain dyadic subdivision of R we have $\mu g \sigma = \nu h \sigma$; $\mu g B d\sigma = \nu h B d \sigma = 0$;* diam $g\sigma < \epsilon$, diam $h\sigma < \epsilon$, diam $\sigma < \epsilon$.

By Lemma 9, there exists an automorphism g_1 such that for every dyadic cell σ we have $\mu g_1 B d\sigma = 0$. Let μ_1 be the measure $\mu_1 A = \mu g_1 A$. Let $\sigma_1, \cdots, \sigma_N$ be the cells of a dyadic subdivision such that diam $\sigma_i < \epsilon$ and diam $g_1\sigma_i < \epsilon$, $i = 1, \cdots, N$. The numbers $\mu_1\sigma_i$ are all positive and their sum is equal to νR. Hence, by Lemma 8, there exists an automorphism h_1 such that $\nu h_1 \sigma_i = \mu_1 \sigma_i$, $i = 1, \cdots, N$. Put $\nu_1 A = \nu h_1 A$. Then since $\nu_1 B d \sigma_i = 0$, we can apply Lemma 9 to each cell σ_i and get an automorphism h_2 that transforms each cell σ_i into itself, such that $\nu_1 h_2 B d\sigma = 0$ for every dyadic cell σ. Let σ_{ij}, $i = 1, \cdots, N$, $j = 1, \cdots, M$, $\sigma_{ij} \subset \sigma_i$, be the cells of a finer dyadic subdivision of R, such that diam $h_1 h_2 \sigma_{ij} < \epsilon$. Since the numbers $\nu_1 h_2 \sigma_{ij}$ are all positive, and $\sum_j \nu_1 h_2 \sigma_{ij} = \mu_1 \sigma_i$, we can apply Lemma 8 to each cell σ_i using the measure μ_1 and get an automorphism g_2 that transforms each cell σ_i into itself, such that $\mu_1 g_2 \sigma_{ij} = \nu_1 h_2 \sigma_{ij}$. That is, $\mu g_1 g_2 \sigma_{ij} = \nu h_1 h_2 \sigma_{ij}$. Since the sum of these MN numbers is equal to μR (or νR), we have also $\mu g_1 g_2 B d \sigma_{ij} = \nu h_1 h_2 B d \sigma_{ij} = 0$. Finally, since $h_1 h_2 \sigma_{ij} \subset h_1 \sigma_{ij}$, $g_1 g_2 \sigma_{ij} \subset g_1 \sigma_{ij}$, and $\sigma_{ij} \subset \sigma_i$, we have also diam $h_1 h_2 \sigma_{ij} < \epsilon$, diam $g_1 g_2 \sigma_{ij} < \epsilon$, and diam $\sigma_{ij} < \epsilon$, and so the automorphisms $g = g_1 g_2$, $h = h_1 h_2$ fulfill all the requirements of the lemma.

LEMMA 11: *Any two r-dimensional Lebesgue-Stieltjes measures μ, ν in R such*

[30] Cf. footnote 28.

that $\mu R = \nu R$ are automorphic to each other under an automorphism that leaves the boundary fixed.

The proof consists in finding a sequence of partitions of R into cells[31] $\sigma_1^{(n)}, \cdots, \sigma_{N_n}^{(n)}$ and two sequences of automorphisms g_n, $h_n \in H_0[R]$, with the following properties, where $\epsilon_n = \frac{1}{2^n} \operatorname{diam} R$.

1° $\operatorname{diam} \sigma_i^{(n)} \leqq \epsilon_n$, $\operatorname{diam} g_n\sigma_i^{(n)} \leqq \epsilon_n$, $\operatorname{diam} h_n\sigma_i^{(n)} \leqq \epsilon_n$

2° $g_n\sigma_i^{(n-1)} = g_{n-1}\sigma_i^{(n-1)}$, $h_n\sigma_i^{(n-1)} = h_{n-1}\sigma_i^{(n-1)}$, $i = 1, \cdots, N_{n-1}$

3° $\mu g_n Bd\sigma_i^{(n)} = \nu h_n Bd\sigma_i^{(n)} = 0$, $i = 1, \cdots, N_n$

4° $\mu g_n\sigma_i^{(n)} = \nu h_n\sigma_i^{(n)}$, $i = 1, \cdots, N_n$

These conditions are satisfied for $n = 0$ if we take g_0, h_0 equal to the identity, and $\sigma^{(0)} = R$. Suppose g_n, h_n; $\sigma_1^{(n)}, \cdots, \sigma_{N_n}^{(n)}$ have been defined so that conditions 1° to 4° are satisfied. Let $\mu_n A = \mu g_n A$ and $\nu_n A = \nu h_n A$. Then for each cell $\sigma_i^{(n)}$ we have $\mu_n\sigma_i^{(n)} = \nu_n\sigma_i^{(n)}$ and $\mu_n Bd\sigma_i^{(n)} = \nu_n Bd\sigma_i^{(n)} = 0$. Hence μ_n and ν_n satisfy the conditions of Lemma 10 relative to $\sigma_i^{(n)}$, and there exist automorphisms g', $h' \in H_0[\sigma_i^{(n)}]$ such that for each cell $\sigma_{ij}^{(n)}$ of a certain dyadic subdivision of $\sigma_i^{(n)}$ we have $\mu_n g'\sigma_{ij}^{(n)} = \nu_n h'\sigma_{ij}^{(n)}$; $\mu_n g' Bd\sigma_{ij}^{(n)} = \nu_n h' Bd\sigma_{ij}^{(n)} = 0$; and $\operatorname{diam} g'\sigma_{ij}^{(n)} < \epsilon$, $\operatorname{diam} h'\sigma_{ij}^{(n)} < \epsilon$, $\operatorname{diam} \sigma_{ij}^{(n)} < \epsilon$, where we suppose ϵ chosen less than ϵ_{n+1}, and so small that both g_n and h_n take sets of diameter less than ϵ into sets of diameter less than ϵ_{n+1}. Then we have $\mu g_n g'\sigma_{ij}^{(n)} = \nu h_n h'\sigma_{ij}^{(n)}$; $\mu g_n g' Bd\sigma_{ij}^{(n)} = \nu h_n h' Bd\sigma_{ij}^{(n)} = 0$; and $\operatorname{diam} g_n g'\sigma_{ij}^{(n)} < \epsilon_{n+1}$, $\operatorname{diam} h_n h'\sigma_{ij}^{(n)} < \epsilon_{n+1}$, $\operatorname{diam} \sigma_{ij}^{(n)} < \epsilon_{n+1}$. The automorphisms g', h' thus defined in each of the cells $\sigma_i^{(n)}$ join up to form automorphisms of R, which we shall also denote by g', h'. Putting $g_{n+1} = g_n g'$, $h_{n+1} = h_n h'$ and taking the totality of cells $\sigma_{ij}^{(n)}$ for $\sigma_1^{(n+1)}$, $\cdots, \sigma_{N_{n+1}}^{(n+1)}$, we see that conditions 1° to 4° are again satisfied, and the inductive definition is complete.

From conditions 1° to 4° it follows that the sequences g_n, h_n converge in $H_0[R]$ to automorphisms g, h. Because, if $x \in \sigma_i^{(n)}$ then $g_{n+1}x$ and $g_n x$ both belong to $g_n\sigma_i^{(n)}$, by 2°. Hence $| g_{n+1}x - g_n x | < \epsilon_n$,[32] by 1°. Similarly, if $x \in g_{n+1}\sigma_i^{(n)}$ then $g_n^{-1}x$ and $g_{n+1}^{-1}x$ both belong to $\sigma_i^{(n)}$. Hence $| g_{n+1}^{-1}x - g_n^{-1}x | < \epsilon_n$. Therefore $\rho(g_{n+1}, g_n) < \epsilon_n$, and because of our choice of ϵ_n, $\rho(g_{n+k}, g_n) < \epsilon_{n-1}$. Hence the sequence g_n converges in $H_0[R]$, and the sequence h_n for similar reasons. From 2° it follows that $g\sigma_i^{(n)} = g_n\sigma_i^{(n)}$ and $h\sigma_i^{(n)} = h_n\sigma_i^{(n)}$, and therefore that $gBd\sigma_i^{(n)} = g_n Bd\sigma_i^{(n)}$ and $hBd\sigma_i^{(n)} = h_n Bd\sigma_i^{(n)}$. From 3° and 4° we then find that $\mu g\sigma_i^{(n)} = \nu h\sigma_i^{(n)}$ and $\mu g Bd\sigma_i^{(n)} = \nu h Bd\sigma_i^{(n)} = 0$. Hence the measures $\mu' A = \mu g A$ and $\nu' A = \nu h A$ are equal for all cells $\sigma_i^{(n)}$ and are zero for their boundaries. Now from 1° it follows that any open set G can be represented as the union of a sequence of non-overlapping cells $\sigma_i^{(n)}$, say $G = \sum_{s=1}^{\infty} \sigma_s$. Then, since the intersection of any two cells σ_s has measure zero with respect to μ'

[31] The cells will all be dyadic cells, but not necessarily of the same subdivision.

[32] We use the vector notation $| x - y |$ to denote the euclidean distance between points x and y.

and ν', we have $\mu' G = \sum_{s=1}^{\infty} \mu' \sigma_s = \sum_{s=1}^{\infty} \nu' \sigma_s = \nu' G$. Hence μ'^* and ν'^* agree for all open sets, and therefore for all sets. Thus $\mu^* A = \nu^* h g^{-1} A$ for every set, and the lemma is proved.

Theorem 2 follows at once on taking $\nu = m$.

In the proof of Lemma 7 it was remarked that in that case the automorphism could be chosen so as to involve only a single coordinate. A simple refinement of the proof of Lemma 9 shows that the automorphism there obtained may be taken to be a finite product of automorphisms each of which involves only a single coordinate. The final automorphism in Lemma 11 is obtained by composing such automorphisms and then passing to the limit. Thus it is possible to assert that the correspondence of the two measure functions can be effected by an automorphism that can be approximated uniformly by finite products of automorphisms each of which involves only a single coordinate. This remark is of interest in view of the fact that it is still an open question whether an arbitrary automorphism of R can be so approximated.[33] If true, this would furnish a strong inductive method for proving topological theorems.

Since the transformation can always be effected by an automorphism that is special at least to the extent of leaving the boundary fixed, and perhaps in other respects, as indicated in the last paragraph, it is natural to inquire whether it would suffice to consider only differentiable automorphisms, or ones whose modulus of continuity is otherwise restricted. That this is not the case may be seen as follows.

Consider the r-dimensional unit cube and let h be a radial contraction that leaves the boundary fixed and takes each concentric cube with edge length $2d < \frac{1}{2}$ into the one whose circumscribed sphere has radius $\delta = e^{-\frac{1}{d}}$, and let $\mu A = m h^{-1} A$. Then the μ-measure of any sphere about the center with radius $\delta < e^{-4}$ is greater than the Lebesgue measure of a sphere with radius $d = \left(\log \frac{1}{\delta}\right)^{-1}$. Hence any automorphism that takes μ into m must have a modulus of continuity at the center greater than $\left(\log \frac{1}{\delta}\right)^{-1}$. But this function dominates any of the form $C\delta^{\alpha}$, $\alpha > 0$, and so the automorphism cannot be differentiable at the center, nor can it satisfy even a Lipschitz or Hölder condition there.

III. Proof of Theorem 1

In Part I, the proof of Theorem 1, as regards category, was reduced to Lemma 5, except for an assumption in the proof of Lemma 3 which has now been removed. Lemma 5 is an approximation theorem. To prove it, it will be necessary to devise methods for modifying a given transformation so as to secure control over the distribution of the images of a certain closed set. We shall do

[33] Problem proposed by one of us in Fund. Math. **24** (1935) 324.

this in three main steps. In §5 it is shown how a transformation can be modified so that it will have a periodic point with images equally distributed among the cells of a given dyadic subdivision. In §6 a lemma is derived which, in effect, enables one to obtain a periodic Cantor set, starting with only a periodic point. These two constructions are combined in §7, and the final step consists in an equivalence transformation which expands the Cantor set to one having the required amount of measure. The result is that an arbitrary measure-preserving automorphism of a rectangular r-cell (or, more generally, a transformation in the r-cell that corresponds to a measure-preserving automorphism of another polyhedron as described in Lemma 4) can be modified so that the resulting transformation is measure-preserving and has a periodic Cantor set whose images carry a prescribed fraction of the total measure and are equally distributed among the cells of a given dyadic subdivision. All modifications are effected by left and right multiplication with automorphisms that leave the boundary fixed. Thus Lemma 5 is established, and therefore Theorem 1. In §18 it will be shown that for automorphisms in $M_0[R]$ one can obtain still more precise approximation theorems.

5. Lemmas Concerning Distribution of Finite Sets

LEMMA 12: *Let p, q be any two interior points of a rectangular r-cell R, $r \geqq 2$. There exists an automorphism $T \in M_0[R]$ such that $Tp = q$.*

Observe first that there is no difficulty in defining an automorphism g that carries p into q and leaves BdR fixed. Most simply, join p to the vertices of R and map the resulting cell complex affinely onto the similar complex obtained from q. Thus the only point to the lemma is to show that the transformation can be effected by a measure-preserving automorphism. It is possible to do this directly by a construction involving displacements around closed tubes, but a more elegant proof is based on Theorem 2. Let g be the automorphism just defined, and consider the outer measure $\mu^*A = m^*g^{-1}A$. By Theorem 2_1 there exists an automorphism $h \in H_0[R]$ such that $\mu^*A = m^*hA$. By Corollary 4, h can be so chosen as to leave q fixed, in fact it may leave a line segment through q fixed. The composed automorphism hg then carries p into q, leaves BdR fixed, and since $m^*hgA = \mu^*gA = m^*A$ we see that $hg \in M_0[R]$.

LEMMA 13: *Given two sets of points $p_1, \cdots, p_N$; $q_1, \cdots, q_N$, each consisting of N distinct interior points of a rectangular r-cell R, $r \geqq 2$, and $|p_i - q_i| < \epsilon$, there exists an automorphism $T \in M_0[R]$ such that $Tp_i = q_i$, $i = 1, \cdots, N$, and $\rho(T, I) < \epsilon$.*

In effect, the lemma asserts that any two finite sets of interior points that could possibly be equivalent under automorphism are equivalent under a measure-preserving automorphism whose norm is no larger than it need be. Let L_i denote the straight line segment joining p_i to q_i. Then L_i is contained in the interior of R. Suppose first that these segments are all disjoint. Then we can enclose each in the interior of a rectangular r-cell σ_i so chosen that the cells $\sigma_1, \cdots, \sigma_N$ are disjoint and each has diameter less than ϵ. By the previous

lemma, we can find an automorphism T_i of each cell σ_i which preserves measure and carries p_i into q_i. Since these automorphisms leave the boundary of each cell σ_i fixed, they can be extended to an automorphism T of R by defining T equal to T_i in each cell σ_i and equal to the identity elsewhere. Since T only permutes points within each cell, it moves no point by more than ϵ, and so $\rho(T, I) < \epsilon$. In case any of the segments L_i intersect, we shall join each p_i to q_i by a chain of $k + 1$ points $p_i = p_{i0}, p_{i1}, \cdots, p_{ik} = q_i$ in such a way that $| p_{i,j} - p_{i,j+1} | < \epsilon/k$ and such that for each $1 \leqq j \leqq k$ the segment L_{ij} joining $p_{i,j-1}$ to $p_{i,j}$ are disjoint. This can always be done; in fact, unless two of the segments L_i intersect in a point that divides each in the same ratio, it will suffice to take the points p_{ij} as the points that divide L_i into k equal segments. In the exceptional case a slight displacement of some of the intermediate points will make the segments $L_{1j}, \cdots, L_{Nj}$ disjoint and will not disturb the inequalities $| L_{ij} | < \epsilon/k$. By the argument just given above, we can find an automorphism T_j of norm less than ϵ/k such that $T_j p_{i,j-1} = p_{i,j}$, $i = 1, \cdots, N$. Hence the composed transformation $T = T_k T_{k-1} \cdots T_1$ will carry each p_i into q_i. Furthermore, it will have norm less than ϵ, because the product of k transformations whose norms are less than ϵ/k always has norm less than ϵ, as may be deduced at once from the triangle inequality.

LEMMA 14: *In a rectangular r-cell R, $r \geqq 2$, let D be a set with measure equal to mR. Let U_δ denote the operation of "taking the δ-neighborhood with respect to D," so that $U_\delta A$ means the set of all points in D at distance less than δ from A. To each $\delta > 0$ corresponds a positive integer λ with the following property: If T is any 1:1 measure-preserving transformation of D onto itself and p is any point of D then $(TU_\delta)^\lambda p = D$.*

It is to be emphasized that λ depends only upon δ and the dimension number r, is independent of the transformation T and of the set D. The transformation need not be assumed to be an automorphism, although it will be in the only application we shall make.

Essentially, the proof consists in the observation that the composite operation TU_δ operating on any non-void set either increases its measure by at least a certain minimum amount or else yields the whole domain D. The important thing is to obtain a uniform estimate of this minimum amount.

The integer λ is determined as follows. Let η be the minimum volume of the intersection of the δ-neighborhoods of any two points of R at distance δ apart. This number η is not simply the common volume of two such spheres, because when the two points are near the boundary of R part of the intersection of their δ-spheres may fall outside R. However, we are concerned only with the fact that $\eta > 0$, and a lower bound to η is furnished by the number $V/2^r$, where V is the r-dimensional volume of a sphere of radius $\delta/2$. This is because the intersection of the δ-neighborhoods of any two points at distance δ certainly contains the $\delta/2$-neighborhood of the point of R midway between them, and the minimum volume of the $\delta/2$-neighborhood of any point of R is $V/2^r$, that being the volume of the $\delta/2$-neighborhood of a vertex of R. We shall take for λ the least integer greater than mR/η.

Now consider the composite operation TU_δ. We shall show that λ iterations of this operation, starting from any point of D, yields the whole of D. For let us suppose that $(TU_\delta)^\lambda p$ is not equal to D, and consider any integer n, $0 \leqq n < \lambda$. The set $(TU_\delta)^n p$ is then not δ-dense in D, and so there is a point q in R whose distance from it is exactly δ. Let q' be a limit point of the set $(TU_\delta)^n p$ at distance δ from q. The operation U_δ applied to the set $(TU_\delta)^n p$ therefore adjoins at least the common part in D of the δ-neighborhoods of q and q', and therefore increases the measure of the set by at least η. That is, $mU_\delta(TU_\delta)^n p \geqq m(TU_\delta)^n p + \eta$. Hence, recalling that T preserves measure, $m(TU_\delta)^{n+1} p \geqq m(TU_\delta)^n p + \eta$. This being true for $n = 0, 1, \cdots, \lambda - 1$, it follows that $m(TU_\delta)^\lambda p \geqq \lambda\eta$, which is impossible, since by definition $\lambda\eta > mR$.

LEMMA 15: *Let T be any measure-preserving automorphism of a set D contained in the interior of an r-cell R, $r \geqq 2$, with $mD = mR$. There exist arbitrarily small measure-preserving automorphisms S_1, $S_2 \in M_0[R]$ such that under the transformation $S_2S_1TS_2^{-1}$ the centers of the cells of a certain arbitrarily fine dyadic subdivision of R undergo a cyclic permutation.*

This lemma is in a sense the key to the present problem. It enables one to gain control over the successive images of a certain point, in fact, it secures their equal distribution among the cells of a certain dyadic subdivision. Since the proof is somewhat involved, it may be well to sketch the idea in advance. It is first shown that one can find points $p_1, \cdots, p_K$ whose first L images under T constitute a set which is "nearly" uniformly distributed among the cells of a given dyadic subdivision. This much is deduced from the ergodic theorem. (This is the only place in the entire proof of Theorem 1 where the ergodic theorem of Birkhoff is used.) But some of the points of this set may coincide. The next part of the proof consists in modifying T by composition with a small transformation S_1 to secure that a similarly distributed set of points are all distinct and *also* are linked together to form a single cycle. In joining the chains together, it is necessary to add more points, but it is secured that only a "few" points are added. The result is that under S_1T there is a cycle consisting of points $q_1, \cdots, q_s$ of which KL are distributed exactly like the first L images of $p_1, \cdots, p_K$. It is further secured that the number s is equal to the number of cells in a finer subdivision, and that the points $q_1, \cdots, q_s$ are sufficiently nearly uniformly distributed that they can be set in correspondence with the centers of the cells of this finer subdivision in such a way that corresponding points are close together. By a final transformation S_2 the points $q_1, \cdots, q_s$ are carried into these centers, so that the latter constitute a cycle under $S_2S_1TS_2^{-1}$.

Proceeding now to a more precise formulation of this proof, let δ be an arbitrary positive number, and let $\sigma_1, \cdots, \sigma_N$ be the cells of a dyadic subdivision with diameter less than δ, say $N = 2^{\alpha r}$. Let them be numbered in such a way that σ_i has at least one vertex in common with σ_{i+1}, $1 \leqq i < N$. This is easily seen to be possible. Let $f_i(p)$ be the characteristic function of the interior of σ_i. The ergodic theorem asserts that the limit $f_i^*(p) = \lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} f_i(T^\nu p)$

exists for almost all p, and that $\int_D f_i^*(p)\,dm(p) = \int_D f_i(p)\,dm(p) = 1/N$, provided we suppose that $mR = 1$. We shall show first that we can select a finite set of points $p_1, \cdots, p_K$ such that the average of each of the functions $f_i^*(p)$ over this set is nearly equal to its integral, that is, $\left|\frac{1}{K}\sum_{k=1}^{K} f_i^*(p_k) - \frac{1}{N}\right| < \eta$. Here η may be any positive number, but for our purpose it will suffice to take $\eta = 1/N^2$. Let J be an integer greater than $2/\eta$, and for each set of positive integers $i_1, \cdots, i_N$ not exceeding J let $A_{i_1 \cdots i_N}$ denote the set of points p in D such that $\frac{i_j - 1}{J} < f_j^*(p) \leqq \frac{i_j}{J}$ for all integers $1 \leqq j \leqq N$. Let $p_{i_1 \cdots i_N}$ be any point of $A_{i_1 \cdots i_N}$, provided this set has positive measure, otherwise undefined. Then, since on $A_{i_1 \cdots i_N}$ the functions $f_j^*(p)$ differ from $f_j^*(p_{i_1 \cdots i_N})$ by not more than $1/J$, we have $|1/N - \sum' f_j^*(p_{i_1 \cdots i_N}) \cdot mA_{i_1 \cdots i_N}| \leqq 1/J < \eta/2$, where $\sum'$ denotes summation over all sets of indices $i_1, \cdots, i_N$ such that $mA_{i_1 \cdots i_N} > 0$. Now let $r_{i_1 \cdots i_N}$ be rational numbers such that $|1/N - \sum' f_j^*(p_{i_1 \cdots i_N}) r_{i_1 \cdots i_N}| < \eta/2$ and $\sum' r_{i_1 \cdots i_N} = 1$. Such numbers exist, since $\sum' mA_{i_1 \cdots i_N} = mR = 1$. Let K be a common denominator of all the numbers $r_{i_1 \cdots i_N}$, and select $Kr_{i_1 \cdots i_N}$ points from each set $A_{i_1 \cdots i_N}$ that has positive measure. Call these points $p_1, \cdots, p_K$. Then we have

$$\left|\sum{}' f_j^*(p_{i_1 \cdots i_N}) r_{i_1 \cdots i_N} - \frac{1}{K}\sum_{k=1}^{K} f_j^*(p_k)\right| \leqq \frac{1}{J} < \eta/2$$

and therefore $|1/N - 1/K \sum_{k=1}^{K} f_j^*(p_k)| < \eta$, $j = 1, \cdots, N$. We may suppose the points $p_1, \cdots, p_K$ so chosen that no image falls on the boundary of a cell σ_j, since this means avoiding only a set of measure zero in making the selection. Recalling that $f_j^*(p_k) = \lim_{n \to \infty} \frac{1}{n} \sum_{\nu=1}^{n} f_j(T^\nu p_k)$, we see that for all sufficiently large n we have $\left|1/N - 1/K \sum_{k=1}^{K} \frac{1}{n} \sum_{\nu=1}^{n} f_j(T^\nu p_k)\right| < \eta$, $j = 1, \cdots, N$. Hence for all sufficiently large n we have also

$$\left|\frac{1}{N} - \frac{1}{Kn}\sum_{k=1}^{K}\sum_{\nu=1}^{n} f_j(T^\nu p_k)\right| + \frac{2(\lambda + 1)}{n} < \eta \tag{1}$$

where λ is the integer corresponding to δ determined in Lemma 14. Let L be such a value of n, and in addition let it be so chosen that for some integer $\beta > \alpha$ we have $2^{\beta r} - K \leqq K(L + \lambda) < 2^{\beta r}$. Such a number L can be found since the numbers $K(n + \lambda)$ include all multiples of K greater than some number. If we let $s = 2^{\beta r}$ and $K_1 = s - K(L + \lambda)$, then $0 \leqq K_1 \leqq K$.

We now proceed to select s distinct points $q_1, \cdots, q_s$ from D in such a way that Tq_i is within distance δ of q_{i+1}, $1 \leqq i \leqq s$, (letting $q_{s+1} = q_1$), and such that KL of these points are interior to the same cells σ_j as corresponding points

$T^\nu p_k$, $1 \leqq \nu \leqq L$, $1 \leqq k \leqq K$. To secure this, we proceed step by step as follows. Take $q_1 = Tp_1$. For $1 < i \leqq L$, take q_i in the same cell as $T^i p_1$ and such that Tq_i is in the same cell as $T^{i+1}p_1$. By Lemma 14 applied to T^{-1}, $q_L \epsilon (T^{-1}U_\delta)^\lambda p_2$. Hence $Tq_L \epsilon U_\delta(T^{-1}U_\delta)^{\lambda-1}p_2$. Choose q_{L+1} in $(T^{-1}U_\delta)^{\lambda-1}p_2$ within distance δ of Tq_L. For $1 < i \leqq \lambda - 1$, take $q_{L+i} \epsilon (T^{-1}U_\delta)^{\lambda-i}p_2$ within distance δ of Tq_{L+i-1}. Then $Tq_{L+\lambda-1}$ is within distance δ of p_2. Take $q_{L+\lambda}$ within distance δ of $Tq_{L+\lambda-1}$ and such that $Tq_{L+\lambda}$ is in the same cell as Tp_2. For $1 \leqq i \leqq L$, take $q_{L+\lambda+i}$ in the same cell as $T^i p_2$ and such that $Tq_{L+\lambda+i}$ is in the same cell as $T^{i+1}p_2$. The next λ points are then chosen to lead up to p_3, similarly to the way in which q_{L+i}, $1 \leqq i \leqq \lambda$, were chosen leading up to p_2. Eventually, the point $q_{(K-1)(L+\lambda)+L}$ is chosen in the same cell as $T^L p_K$. We then close the cycle by a chain of $\lambda + K_1$ points (instead of λ points as in the preceding cases) in order to get a cycle with exactly s points. This can be done since $q_{(K-1)(L+\lambda)+L} \epsilon (T^{-1}U_\delta)^{\lambda+K_1}p_1$. At the last step we have $q_{s-1} \epsilon T^{-1}U_\delta p_1$ and choose q_s near enough to p_1 so that it is within distance δ of Tq_{s-1} and so that Tq_s is within distance δ of $q_1 = Tp_1$. At every step the point to be selected can be chosen arbitrarily from an open sphere with respect to D. Hence there is no difficulty in avoiding points previously chosen and thus securing that the points $q_1, \cdots, q_s$ are distinct, and therefore also the points $Tq_1, \cdots, Tq_s$. We see that the KL points $q_1, q_2, \cdots, q_L, q_{L+\lambda+1}, \cdots, q_{L+\lambda+L}, \cdots, q_{(K-1)(L+\lambda)+1}, \cdots, q_{(K-1)(L+\lambda)+L}$ are respectively interior to the same cells σ_j as the points Tp_1, $T^2p_1, \cdots, T^Lp_1, Tp_2, \cdots, T^Lp_2, \cdots, Tp_K, \cdots, T^Lp_K$. Finally, it is seen that q_{i+1} is always within distance δ of Tq_i, since they either lie in the same cell, or were specifically chosen so as to be within this distance.

In virtue of Lemma 13, we can find an automorphism $S_1 \epsilon M_0[R]$, with norm less than δ, which carries Tq_i into q_{i+1}, $1 \leqq i \leqq s$. Under the composed transformation S_1T the points $q_1, \cdots, q_s$ thus form a single cycle. Furthermore, since KL of these lie in the same cells as corresponding points $T^\nu p_k$, it follows that

$$\left|\frac{1}{N} - \frac{1}{s}\sum_{i=1}^{s} f_j(q_i)\right| \leqq \left|\frac{1}{N} - \frac{1}{s}\sum_{k=1}^{K}\sum_{\nu=1}^{L} f_j(T^\nu p_k)\right| + \frac{K\lambda + K_1}{s}$$

$$\leqq \left|\frac{1}{N} - \frac{1}{KL}\sum_{k=1}^{K}\sum_{\nu=1}^{L} f_j(T^\nu p_k)\right| + \frac{s - KL}{s} + \frac{K\lambda + K_1}{s}.$$

From equation (1) it follows that this is less than η, because

$$\frac{s - KL}{s} + \frac{K\lambda + K_1}{s} = \frac{2(K\lambda + K_1)}{s} \leqq \frac{2K(\lambda + 1)}{K(L + \lambda) + K_1} < \frac{2(\lambda + 1)}{L}.$$

Hence the relative number of points $q_1, \cdots, q_s$ in any cell σ_j differs from $1/N$ by less than $\eta = 1/N^2$. Let us now re-number the points $q_1, \cdots, q_s$ counting first all those in σ_1, then those in σ_2, and so on. Call the resulting sequence $q_1', q_2', \cdots, q_s'$. Let $q_1'', q_2'', \cdots, q_s''$ be the centers of the $s = 2^{\beta r}$ cells of the β-th dyadic subdivision, again counting first those in σ_1, then those in σ_2, and

so on. We shall show that corresponding points q_i' and q_i'' always lie either in the same cell σ_j or in adjacent cells. The number of points q_i in any cell σ_j differs from s/N by less than s/N^2. Hence the indices of the points q_i' that lie in σ_j fall between the limits $\frac{(j-1)s}{N} - \frac{js}{N^2}$ and $\frac{js}{N} + \frac{js}{N^2}$, which are contained between $\frac{(j-2)s}{N}$ and $\frac{(j+1)s}{N}$, the limits of the indices of points q_i'' lying in σ_{j-1}, σ_j, and σ_{j+1}. Since these are adjacent cells, the distance from q_i' to q_i'' is less than 2δ, and by Lemma 13 we can find an automorphism $S_2 \in M_0[R]$ with norm less than 2δ which carries q_i' into q_i'', $1 \leqq i \leqq s$. The points q_i'' then constitute a single cycle under the transformation $S_2S_1TS_2^{-1}$. Since S_1 and S_2 are both arbitrarily near the identity, the lemma is established.

6. Equivalence of Cantor sets under automorphism of the containing space

LEMMA 16: *Let C be a linear Cantor set*[34] *contained in the interior of an r-cell σ_0, which in turn is contained in an r-cell R, $r \geqq 2$, and let f be a homeomorphism of σ_0 onto a subset of R. There exists an automorphism h of R that leaves the boundary fixed and carries fC back into coincidence with C in such a way that $hfp = p$ for every p in C.*

The lemma implies that any topological Cantor set C' contained in the interior of R is equivalent under automorphism of R to a linear Cantor set C provided some homeomorphism of C onto C' can be extended to an r-cell containing C. It should be remarked that such an extension is not always possible even when one of the sets is linear, as is shown by Antoine's example in the 3-dimensional cube (See footnote 15).

The first step in the proof consists in showing that if σ_1 and σ_2 are disjoint r-cells contained in the interior of σ_0, there exists an automorphism $h \in H_0[R]$ such that $hf\sigma_1 \subset \sigma_1$ and $hf\sigma_2 \subset \sigma_2$. This may be seen as follows. Let p_1, p_2 be the centers of σ_1, σ_2 respectively. Then fp_1, fp_2 are distinct interior points of R, and we can find an automorphism $h_1 \in H_0[R]$ which carries them back into p_1, p_2 respectively. Then h_1f is a homeomorphism of σ_0 onto a subset of R which leaves p_1, p_2 fixed. By continuity, there exist r-cells σ_1', σ_2' about p_1, p_2 such that $h_1f\sigma_1' \subset \sigma_1$ and $h_1f\sigma_2' \subset \sigma_2$. Let $g \in H_0[\sigma_0]$ be such that $g\sigma_1 \subset \sigma_1'$ and $g\sigma_2 \subset \sigma_2'$. Such an automorphism is easily defined by dividing σ_0 into two cells containing σ_1, σ_2 in their interiors and then drawing the interior points of each part toward p_1, p_2 respectively. The homeomorphism h_1fg then carries σ_1 and σ_2 into subsets of themselves. Now consider the transformation h_2 of R equal to fgf^{-1} in $f\sigma_0$ and equal to the identity in the rest of R. That h_2 is an automorphism of R follows from Brouwer's theorem on invariance of region.[35] According to this theorem, $f\sigma_0$ is a closed region, that is, the closure of a connected open set, and interior and boundary points of σ_0 correspond respectively to

[34] That is, a topological Cantor set contained in a straight line segment.
[35] See e.g. AH p. 396.

interior and boundary points of $f\sigma_0$. Hence fgf^{-1} is an automorphism of $f\sigma_0$ that leaves all boundary points fixed, and can therefore be extended to the rest of R by defining it equal to the identity outside. The automorphism $h = h_1h_2$ belongs to $H_0[R]$ and hf carries σ_1 and σ_2 into subsets of themselves, as required.

Now let us choose a family of r-cells $\sigma_{i_1 \cdots i_n}$; $i_1, \cdots, i_n = 1, 2; n = 1, 2, \cdots$; contained in the interior of σ_0, so as to fulfill the following conditions:

1° For each n, the cells $\sigma_{i_1 \cdots i_n}$ are disjoint.

2° $\sigma_{i_1 \cdots i_n}$ is contained in the interior of $\sigma_{i_1 \cdots i_{n-1}}$.

3° The diameter of $\sigma_{i_1 \cdots i_n}$ tends to zero with $1/n$.

4° $C = \prod_{n=1}^{\infty} \sum_{i_1 \cdots i_n} \sigma_{i_1 \cdots i_n}$.

Such a representation of any linear Cantor set is immediate from its definition. We have shown that there exists an automorphism $h_1 \in H_0[R]$ such that $h_1 f\sigma_1 \subset \sigma_1$ and $h_1 f\sigma_2 \subset \sigma_2$. Similarly, let h_2 be an automorphism of R that is equal to the identity outside σ_1 and σ_2 and such that $h_2h_1f\sigma_{i_1 i_2} \subset \sigma_{i_1 i_2}$; $i_1, i_2 = 1, 2$. This involves only an application of the previous result to the two cells σ_1, σ_2 separately. At the n-th stage, we find an automorphism $h_n \in H_0[R]$ which is equal to the identity outside the cells $\sigma_{i_1 \cdots i_{n-1}}$ and is such that $h_nh_{n-1} \cdots h_1 f\sigma_{i_1 \cdots i_n} \subset \sigma_{i_1 \cdots i_n}$. The successive products $h_nh_{n-1} \cdots h_1$ converge uniformly, in virtue of condition 3°, and therefore have for limit a continuous mapping h of R onto itself. That h is likewise 1:1 may be seen as follows. Consider two distinct points p, q outside fC. They are outside the sets $f\sigma_{i_1 \cdots i_n}$ for some n. Hence their images under h are the same as under the finite product $h_n \cdots h_1$, and therefore distinct and outside C. On the other hand, h carries distinct points of fC into distinct points of C. In fact, h is equal to f^{-1} on fC, because $hf\sigma_{i_1 \cdots i_n} \subset \sigma_{i_1 \cdots i_n}$ and therefore hf leaves every point of C fixed. Thus h is a 1:1 continuous map of R onto itself, therefore an automorphism, and it leaves boundary points fixed since it is equal to h_1 there. Furthermore, we have just seen that $hfp = p$ for all p in C.

Before proceeding with our main investigation, which will be resumed beginning with Lemma 17, we shall digress to consider the bearing of the present lemma on the work of Antoine.[36] Given two Cantor sets C and C' situated in a euclidean space $E^{(r)}$ three possibilities are conceivable: (i) there is a homeomorphism of C onto C' that can be extended to the whole space; (ii) there is a homeomorphism that can be extended to a neighborhood of C but none can be extended to the whole space; (iii) no homeomorphism can be extended even to a neighborhood of C. Antoine showed that in the plane only case (i) can arise, indeed that the homeomorphism can be taken equal to the identity outside any given rectangle to which C and C' are interior. But he showed by examples that in 3-space (and therefore in r-space, $r > 3$) all three possibilities can arise, and that case (iii) can be realized even when one of the sets is linear. In his example illustrating case (ii), however, both sets are skew. We shall show

[36] See footnote 15.

that this is necessarily the case, that is, that if either of the sets is plane or linear case (ii) cannot arise in any space $E^{(r)}$. More precisely, the result is as follows.

COROLLARY: *If C and C' are Cantor sets contained in $E^{(r)}$, $r \geqq 2$, and C is plane or linear then C and C' are equivalent under automorphism of $E^{(r)}$ if and only if some homeomorphism of C onto C' can be extended to a neighborhood of C. If f is any homeomorphism of C onto C' that can be extended to a neighborhood of C then f can be extended to $E^{(r)}$ in such a way that it is equal to the identity outside any given rectangular r-cell to which C and C' are interior.*

Only the second assertion need be proved since it obviously implies the first. We shall consider first the case in which C is linear. Let R be the given rectangular r-cell containing C and C' in its interior and let f_1 be an extension of f to an open set G containing C. We may suppose that both G and f_1G are contained in R. Using the Heine-Borel theorem we can find a finite number of disjoint rectangular r-cells $R_1, \cdots, R_N$ contained in G whose interiors cover C. Let p_i be the center of R_i. Then $f_1p_1, \cdots, f_1p_N$ are distinct interior points of R and by Lemma 13 there exists $h_1 \in H_0[R]$ such that $h_1f_1p_i = p_i$. Let σ_i' be an r-cell about p_i so small that $h_1f_1\sigma_i' \subset R_i$, and let σ_i be an r-cell interior to R_i and containing CR_i in its interior. Let g be a shrinking transformation of the interior of each cell R_i such that $g\sigma_i \subset \sigma_i'$ and such that g leaves boundary points of $R_1, \cdots, R_N$ fixed. Then h_1f_1g is a homeomorphism of each σ_1 onto a subset of R_i. By Lemma 16 applied to each cell R_i there exists $h_2 \in H_0[R]$ such that $h_2h_1f_1g$ is equal to the identity on C. From the theorem on invariance of region it follows as before that the transformation h_0 equal to $f_1gf_1^{-1}$ in $f_1R_1 + \cdots + f_1R_N$ and equal to the identity elsewhere is an automorphism, and it evidently belongs to $H_0[R]$ since $f_1G \subset R$. Hence $h = h_2h_1h_0$ is in $H_0[R]$ and $hfp = h_2h_1f_1gp = p$ for every p in C. Therefore h^{-1} is an extension of f to R, and it may be defined equal to the identity outside R.

It remains to consider the case in which C is plane but not linear. Let R be a rectangular r-cell containing both C and C' in its interior and let C'' be a linear Cantor set in the same plane as C and likewise interior to R. Antoine has shown that there is an automorphism of this plane that carries C into C'' and is equal to the identity outside R. This automorphism can be extended to $E^{(r)}$ in such a way that it is still equal to the identity outside R. This can be done conveniently by projecting the transformation from two points in R on opposite sides of the plane in which C lies, then projecting again from two points in R on opposite sides of the 3-space in which the transformation is already defined. After $r - 2$ steps the desired extension is obtained. Thus we have an automorphism $h \in H_0[R]$ that carries C into C''. The transformation fh^{-1} is a homeomorphism of C'' onto C' that can be extended to a neighborhood of C''. Since C'' is linear there exists an automorphism $g \in H_0[R]$ equal to fh^{-1} on C'', as we have already shown. The automorphism gh is therefore equal to f on C and it can be defined equal to the identity outside R.

LEMMA 17: *If the given homeomorphism f in Lemma 16 is measure-preserving then the automorphism h can also be taken to be measure-preserving.*

Let h be any automorphism fulfilling the requirements of Lemma 16. In R consider the measure function $\mu A = mh^{-1}A$. Then $\mu C = mh^{-1}C = mfC = 0$, since f is now assumed to be measure-preserving. Let L denote a line segment containing C and contained in the interior of R. It is possible that μL may be positive, but in that case we can displace the points of L not in C by an automorphism $g_1 \in H_0[R]$ so chosen that $\mu g_1 L = 0$ and $g_1 p = p$ for $p \in C$. Then consider the measure $\mu_1 A = \mu g_1 A = mh^{-1}g_1 A$. This is an r-dimensional LS measure and in addition $\mu_1 L = 0$. Hence, by Theorem 2 Corollary 4, there exists an automorphism $g_2 \in H_0[R]$ such that $\mu_1 A = mg_2 A$ and $g_2 p = p$ for $p \in L$. Thus we have $mh^{-1}g_1 A = mg_2 A$ for all $A \subset R$, that is, $mg_2 g_1^{-1} hA = mA$. Hence $g_2 g_1^{-1} h$ is measure-preserving, and since both g_1 and g_2 leave the points of C fixed the measure-preserving automorphism $T = g_2 g_1^{-1} h$ fulfills the requirements of the lemma.

7. Proof of Lemma 5

LEMMA 18: *Let R_0 be the interior of a rectangular r-cell R, $r \geqq 2$, and let T be a measure-preserving homeomorphism of an open set $G \subset R_0$ onto an open set $TG \subset R_0$, where $mG = mR$, and let $\sigma_1, \cdots, \sigma_N$ be the cells of any given dyadic subdivision. There exist arbitrarily small automorphisms $h_1, h_2 \in H_0[R]$ such that $h_1 T h_2$ is a measure-preserving homeomorphism of $h_2^{-1}G$ onto $h_1 TG$, where $mh_2^{-1}G = mR$, and such that $h_1 T h_2$ transforms a certain Cantor set C_1 in the following manner: The points of C_1 are all periodic with the same period; the distinct images of C_1 are disjoint and equally distributed among the cells $\sigma_1, \cdots, \sigma_N$; the images of C_1 contain a prescribed fraction $\alpha < 1$ of the measure of each of the cells σ_i.*

In the proof we shall assume $mR = 1$, which involves no loss of generality. Observe first that T may be considered as a measure-preserving automorphism of the set $D = \prod T^n G$, $-\infty < n < +\infty$, which has measure one, since it is the intersection of countably many sets of measure one. By Lemma 15, there exist arbitrarily small automorphisms $S_1, S_2 \in M_0[R]$ such that $T_1 = S_2 S_1 T S_2^{-1}$ permutes cyclically the centers of the cells $\tau_1, \cdots, \tau_{KN}$ of an arbitrarily fine dyadic subdivision, in particular, finer than the given subdivision $\sigma_1, \cdots, \sigma_N$. Now consider T_1 to have domain $S_2 G$, and let $\tau \subset \tau_1$ be an r-cell about the center of τ_1 so small that its first KN images under T_1 are defined and respectively interior to $\tau_2, \tau_3, \cdots, \tau_{KN}, \tau_1$. Let C be a linear Cantor set interior to τ. Since T_1^{KN} is a measure-preserving homeomorphism of τ onto a subset of τ_1, by Lemma 17 there exists an automorphism $S_3 \in M_0[\tau_1]$, which we define equal to the identity outside τ_1, such that under $T_2 = S_3 T_1$ the set C is brought back to coincidence with itself. Under T_2 the points of C are all periodic with period KN, and there is one image of C in each of the cells $\tau_1, \cdots, \tau_{KN}$. Now introduce a measure μ_1 in C by mapping it onto a linear Cantor set with measure $1/KN$, and extend this measure to the images of C by the transformation T_2. Then μ_1 will be a normalized measure invariant under T_2 such that $\mu_1 \tau_i = 1/KN$, $i = 1, \cdots, KN$. Since T_2 preserves both measures μ_1 and m, it will also preserve the measure $\mu A = (1 - \alpha)mA + \alpha\mu_1 A$. It is easily verified that μ^* fulfills the

conditions of Theorem 2_1 with respect to each of the cells τ_i. Hence there exists an automorphism $h \in H_0[R]$, which transforms each cell τ_i into itself, such that $\mu^*A = m^*hA$ for all $A \subset R$. Hence $T_3 = hT_2h^{-1}$ will preserve Lebesgue measure, and the Cantor set $C_1 = hC$ will be periodic under T_3, will have one image in each cell τ_i, and this image will contain the fraction α of the measure of τ_i. Hence the images of C_1 under T_3 are equally distributed among $\sigma_1, \cdots, \sigma_N$, and contain the fraction α of their measure. Expressing T_3 in terms of T, we have $T_3 = hS_3S_2S_1TS_2^{-1}h^{-1}$. The automorphisms S_1 and S_2 could be taken arbitrarily small, and S_3 and h have norms no greater than the diameter of the cells τ_i. Hence T_3 is of the required form h_1Th_2, where h_1 and h_2 belong to $H_0[R]$ and are arbitrarily small. It should be noted that T_3 is measure-preserving even though h_1 and h_2 are not.

Taking $\alpha = \frac{1}{2}$, it is evident that Lemma 18 implies Lemma 5, in fact the periodicity of the set C_1 is for this purpose superfluous.

8. The Borel class of the set of metrically transitive automorphisms

To complete the proof of Theorem 1, it only remains to show that the set of metrically transitive automorphisms is G_δ in $M[E, \mu]$. This is most easily done by utilizing what amounts to a necessary and sufficient condition for metrical transitivity, rather than the definition itself. Consider the space L_2 of functions quadratically integrable over E. Any automorphism $T \in M[E, \mu]$ induces a unitary transformation $Uf(x) = f(Tx)$ of L_2, as is well-known.[37] Let $f_1, f_2, \cdots$ be a sequence of continuous functions dense in L_2. Let $E(i, j, n)$ be the set of all T in $M[E, \mu]$ such that

$$\int_E \left[\frac{1}{n}\sum_{\nu=0}^{n-1} f_i(T^\nu x) - (f_i, 1)\right]^2 dx < \frac{1}{j},$$

where $(f_i, 1)$ denotes the number $\int_E f_i(x)\,dx$. This set is open in $M[E, \mu]$. For suppose T_k, $k = 1, 2, \cdots$, belongs to the complement of $E(i, j, n)$ and that $\rho(T_k, T) \to 0$ as $k \to \infty$. The integrand converges boundedly to the limit and so T also belongs to the complement. Now, the set M_T of metrically transitive automorphisms is represented by

$$M_T = \prod_{i=1}^{\infty}\prod_{j=1}^{\infty}\sum_{n=1}^{\infty} E(i, j, n).$$

For if T is metrically transitive it will belong to $E(i, j, n)$ for all sufficiently large n, in virtue of the mean ergodic theorem of von Neumann.[38] On the other hand, if T is metrically intransitive there exists an invariant function φ not constant, namely, the characteristic function of an invariant set with measure inter-

[37] B. O. Koopman, *Hamiltonian systems and transformations in Hilbert space.* Proc. Nat. Acad. USA. **17** (1931) 315.

[38] J. von Neumann, *Proof of the quasiergodic hypothesis.* Proc. Nat. Acad. USA. **18** (1932) 70.

mediate between zero and μE. Let d denote the distance in L_2 from φ to the axis of constant functions, and choose $1/j < \frac{1}{4} d^2$. Choose f_i from the sphere of radius $d/2$ about φ. Since φ is a fixed point under U, and U is unitary, all images of f_i under U also lie in the sphere, and therefore also all averages $\frac{1}{n} \sum_{\nu=0}^{n-1} f_i(T^\nu x)$. For no n does T belong to $E(i, j, n)$, because the distance from $\frac{1}{n} \sum_{\nu=0}^{n-1} f_i(T^\nu x)$ to the constant function $(f_i, 1)$ is always at least $d/2$, and therefore the squared distance is greater than $1/j$.

This representation of M_T evidently exhibits it as a G_δ set in $M[E, \mu]$.

IV. Metrically Transitive Flows

9. Existence Theorem

The results concerning metrically transitive *automorphisms* contained in the preceding sections make it possible to set up a general procedure for defining metrically transitive *continuous flows*.[39] By a continuous flow we shall mean a one-parameter group of automorphisms T_λ, $-\infty < \lambda < +\infty$, of a space E such that $T_\lambda x$ is continuous in x and λ, and such that $T_{\lambda+\mu} = T_\lambda T_\mu$ and $T_0 = I$.

Theorem 3: *Let E be a regularly connected polyhedron of dimension $r \geqq 3$, and let μ be any r-dimensional Lebesgue-Stieltjes measure in E. There exists a continuous flow in E which is metrically transitive with respect to μ, and which leaves all singular points of E fixed.*

It is sufficient to define a continuous flow in a rectangular r-cell which leaves the boundary fixed and is metrically transitive with respect to ordinary Lebesgue measure. Because the image of such a flow under the map defined in Lemma 3 will be a continuous flow in E, in virtue of Lemma 4, and metrically transitive with respect to μ.

We first define a flow in a space Q_1 defined as follows. Let B denote the $(r - 1)$-dimensional unit cube and introduce in it a measure $\nu A = \int_A f(p)\, dp$, where the integral is an ordinary $(r - 1)$-dimensional Lebesgue integral and $f(p)$ is a continuous function positive over the interior of B and tending to infinity at the boundary in such a way that $\int_B f(p)\, dp = 1$. Then νA is an $(r - 1)$-dimensional LS measure in B, and by Theorem 1 there exists an automorphism T of B which leaves the boundary fixed and is metrically transitive with respect to ν. Consider the product space of B with the unit interval $0 \leqq \lambda \leqq 1$, and identify points $(p, 1)$ and $(Tp, 0)$. In this space Q_1 define a flow upward along streamlines perpendicular to B, taking the velocity at any point to be $1/f(p)$, where p is the last intersection of the streamline with B. The velocity along a streamline undergoes a discontinuous change when it crosses B. Nevertheless,

[39] This is an adaptation of a standard method. See E. Hopf, *Ergodentheorie*, p. 41.

the flow defined by this velocity field is continuous, and since $1/f(p)$ tends to zero at the boundary, the boundary remains fixed. This flow preserves r-dimensional Lebesgue measure in Q_1. To see this, consider any small cube interior to Q_1 with height h and base σ. Until it crosses B, the segments along the streamlines are rigidly translated, so that its volume is not changed. After it crosses B it has a new cross-section $T\sigma$, and the length along any streamline is now $h \cdot f(p)/f(Tp)$. The volume is therefore

$$\int_{T\sigma} h \frac{f(p)}{f(Tp)}\, dp = \int_{T\sigma} \frac{h}{f(Tp)}\, d\nu(p) = \int_{\sigma} \frac{h}{f(p)}\, d\nu(p) = \int_{\sigma} h\, dp.$$

Thus, the change in velocity exactly compensates for the change in cross-section, and therefore the flow is measure-preserving, since a flow that preserves the measure of small cubes preserves the measure of all sets. To see that it is metrically transitive, consider any invariant measurable set. The set must be cylindrical, and so its intersection A with B is Lebesgue measurable and therefore ν-measurable. Since A is invariant under T its ν-measure is either zero or one, and consequently also its Lebesgue measure. Hence the cylindrical set over it has Lebesgue measure zero or one.

The space Q_1 in which we have just defined a flow is homeomorphic to the r-dimensional tube Q, that is, the product space of B with $(0, 1)$ where points $(p, 0)$ and $(p, 1)$ are identified. To prove this, observe that since T leaves the boundary fixed, it can be joined isotopically to the identity. That is, there exists a continuous family of automorphisms T_λ of B, $0 \leqq \lambda \leqq 1$, such that $T_1 = T$ and T_0 is the identity.[40] The correspondence $(p, \lambda) \rightarrow (T_\lambda p, \lambda)$ is a homeomorphism of Q onto Q_1.

The r-dimensional unit cube R can be represented as the continuous image of Q under a map which is a homeomorphism up to the boundary of Q. This may be done as follows. The region of r-space defined by the inequalities $1 \leqq (x_1^2 + x_2^2)^{\frac{1}{2}} \leqq 2$, $0 \leqq x_i \leqq 1$, $i > 2$, is evidently homeomorphic to Q. The map defined by $x_1' = x_1[(x_1^2 + x_2^2)^{\frac{1}{2}} - 1]$, $x_2' = x_2[(x_1^2 + x_2^2)^{\frac{1}{2}} - 1]$, $x_i' = x_i$, $i > 2$, is continuous and has for range the set $0 \leqq (x_1'^2 + x_2'^2)^{\frac{1}{2}} \leqq 2$, $0 \leqq x_i' \leqq 1$, $i > 2$, which is homeomorphic to R. This map is 1:1 except at points where $(x_1^2 + x_2^2)^{\frac{1}{2}} = 1$, which belong to the boundary of Q. Combining this map with an automorphism of R based on Theorem 2, the map from Q_1 to R can be made measure-preserving. The image in R of the flow already defined in Q_1 is therefore metrically transitive and leaves the boundary of R fixed. From this we can derive a metrically transitive continuous flow in E, as already explained.

10. Most general polyhedra that can support metrically transitive automorphisms or flows

In the present section we consider arbitrary finite polyhedra, not necessarily connected or even homogeneous-dimensional, and seek to characterize those in

[40] J. W. Alexander, *On the deformation of an n-cell.* Proc. Nat. Acad. USA. **9** (1923) 406–407.

which metrically transitive continuous flows or automorphisms are possible. It will be seen that a complete answer is obtained except in the case $r = 2$ for flows.

THEOREM 4: *A finite polyhedron of dimension $r \geqq 3$ can support a metrically transitive continuous flow with respect to any given r-dimensional Lebesgue-Stieltjes measure if and only if its regular points form a connected set.*

If the regular points form a connected set, the closure of this set is a regularly connected polyhedron, and we have already seen how to construct a metrically transitive flow in this part. Since the flow given by Theorem 3 leaves singular points fixed, it can be extended to the rest of the polyhedron, which is a set of measure zero, by defining it equal to the identity there. On the other hand, if the regular points fall into two or more disjoint open sets, these will be invariant under any continuous flow, since if one point of a streamline is regular, all must be. There will therefore necessarily exist disjoint invariant sets with positive measure, so that no metrically transitive, or even topologically transitive, continuous flow is possible.

The case $r = 2$ is not covered by our present method of construction, and it appears likely that further conditions on the polyhedron are necessary in this case. The case $r = 1$ is trivial, the only possibility is for the polyhedron to be homeomorphic to the circumference of a circle, plus possibly some isolated points, the flow being topologically equivalent to a steady rotation of the circle. The details are left to the reader.

THEOREM 5: *A finite polyhedron of dimension $r \geqq 2$ can support a metrically transitive automorphism with respect to any given r-dimensional Lebesgue-Stieltjes measure if and only if its regular components have equal measure and can be permuted cyclically by an automorphism.*

Suppose the conditions satisfied. Let h be an automorphism that permutes the regular components cyclically. By Theorem 2 Corollary 3, we can modify h within each regular component so as to make it measure-preserving. Call the resulting automorphism T. Suppose there are n regular components, and let E_1 be one of them. Then T^n is a measure-preserving automorphism of E_1. By the Corollary to Theorem 1, there exists an arbitrarily small automorphism S of E_1 such that ST^n is a metrically transitive automorphism of E_1, and since S leaves all singular points fixed, it can be extended to the rest of E by defining it equal to the identity outside E_1. Then ST is a metrically transitive automorphism of E, because any invariant set of positive measure must contain almost all points of E_1 and therefore almost all points of every regular component. Conversely, if there exists a metrically transitive automorphism it must permute the regular components cyclically, and consequently they must have equal measure. In fact, these conditions are necessary if there is to exist even a topologically transitive measure-preserving automorphism.

It may be added that in the case $r = 1$ the only polyhedron that can support a metrically transitive automorphism is one that is homeomorphic to a finite number of disjoint circumferences of circles of equal measure, plus possibly some

isolated points. In such a polyhedron a metrically transitive automorphism is easily defined by composing a non-periodic rotation of one of the circles with a cyclic permutation of the circles. The isolated points, if any, may be left fixed. Any other 1-dimensional polyhedron will have a regular component with at least one singular point. This component will be either a line segment or a circle with one singular point. Some power of any given automorphism will leave all singular points fixed and each regular component invariant, since there are only a finite number of each. But a measure-preserving automorphism of a line segment that leaves the ends fixed must be the identity, and a measure-preserving automorphism of a circle that leaves one point fixed must be either the identity or a reflection in a diameter. In none of these cases can the automorphism be metrically transitive, since a power of it will be equal to the identity on some regular component.

Thus we have obtained a characterization of all finite polyhedra that can support metrically transitive automorphisms.

It may be remarked that although metrically transitive automorphisms may exist in polyhedra having more than one regular component, nevertheless Theorem 1 is not true for such polyhedra, because we have just seen that a transitive automorphism must permute the regular components cyclically, and therefore cannot be near the identity. However, the following generalization of Theorem 1 holds. *In the space $M[E, \mu]$ of measure-preserving automorphisms of any polyhedron that can support at least one metrically transitive automorphism, the metrically transitive automorphisms form a residual set with respect to the subspace of automorphisms that permute the regular components cyclically.* One need only observe that in the existence proof given above it was really shown that the metrically transitive automorphisms are dense in this subspace. Since they form a G_δ set with respect to the whole space, they are G_δ with respect to the subspace also, and therefore residual.

11. Transitive automorphisms in transitive flows

In the preceding sections we have made use of a general method whereby a transformation can be used to define a flow in a product space. By this construction a transitive transformation gives rise to a transitive flow. But there is an even simpler relation connecting transformations and flows, namely, the individual transformations that make up a continuous flow are themselves automorphisms of the space. If even one automorphism in the flow is transitive, either metrically or topologically, then the flow is, because by definition a set that is invariant under a flow is invariant under each of the transformations that make up the flow. This might seem to suggest another way of deriving a transitive flow from a transformation, but the method is of little use because in general an automorphism cannot be embedded in a flow.[41] Nevertheless, it is natural

[41] Any transformation embedded in a continuous flow must, for example, have roots of all orders. A simple example of an automorphism which hasn't even a square root is the following. Let T_1 be a rotation of the circumference of a circle through an angle π/k,

to inquire whether any of the automorphisms that make up a transitive continuous flow are necessarily transitive. In the case of metrical transitivity this question appears to be open.[42] It is therefore of interest to note that in the topologically transitive case the answer is in the affirmative, as we proceed to show.

THEOREM 6: *Let T_λ, $-\infty < \lambda < +\infty$, be a topologically transitive continuous flow in a separable metric space E, and suppose there is no isolated streamline. For all values of λ, except a set of first category on the line $-\infty < \lambda < +\infty$, the automorphisms T_λ are topologically transitive.*

Let G_1, G_2, $\cdots$ be a countable base for all open sets in E. Let $A_{i,j}$ be the set of values of λ such that for some positive or negative integer k the set $T_{k\lambda}G_i$ overlaps G_j, that is, such that G_i overlaps G_j under some power of T_λ. The set $A_{i,j}$ is evidently open, we proceed to show that it is dense on the line $-\infty < \lambda < +\infty$. Consider any interval I and form the set of all numbers $k\lambda$, where $\lambda \in I$ and $k = 0, \pm 1, \pm 2, \cdots$. This set includes all numbers greater in absolute value than some number Λ. Form the set $H = \sum_{|\lambda| > \Lambda} T_\lambda G_i$ and consider a sphere $\sigma \subset G$ so small that the set swept out by it as λ describes the interval $-\Lambda \leqq \lambda \leqq \Lambda$ is not dense in G_j. Any sufficiently small sphere about a point of G_i will do since by hypothesis there is no isolated streamline. The set H must overlap G_j, because otherwise the invariant open set swept out by σ would not be dense in G_j, contrary to the hypothesis that the flow is topologically transitive. Hence there exist values of λ greater in absolute value than Λ such that $T_\lambda G_i$ overlaps G_j, and so the set $A_{i,j}$ contains points of the interval I. Since the sets $A_{i,j}$ are open and everywhere dense, their intersection as i and j run independently over all positive integers is a residual set. For any value of λ in this set, the automorphism T_λ is topologically transitive, because if it had an invariant open set not everywhere dense in E there would exist a pair G_i, G_j such that no image of G_i overlaps G_j, and this would contradict the fact that λ belongs to $A_{i,j}$.

V. SOME RELATED QUESTIONS

12. Generation of Random Sequences by Automorphisms

In the Introduction it was explained that metrical transitivity implies that the images of almost all points are distributed like random sequences in respect

---

where k is a positive integer. Let T_2 be the automorphism of the segment $0 \leqq \theta \leqq \frac{\pi}{k}$ defined by $T_2\theta = \left(\frac{k\theta}{\pi}\right)^2$, and let it leave all other points fixed. Under the automorphism $T = T_2T_1$, the points $0, \frac{\pi}{k}, \frac{2\pi}{k}, \cdots, \frac{(2k-1)\pi}{k}$ form a single cycle of period $2k$, and no other points are periodic. It is clear that T cannot be the square of a transformation S, because a cycle of period $2k$ in T could arise only by the splitting of a cycle of period $4k$ in S, and then T would have two cycles of period $2k$. Hence T has no square root.

[42] The question was raised by H. E. Robbins.

to the frequency with which they fall in any given measurable set. A random sequence has also the property that of its first n points the number that fall in A differs from $n \cdot mA$ by something of the order of $\sqrt{n}$. Furthermore, this difference oscillates in sign. Let us consider the r-dimensional unit cube R and take for A any dyadic cell σ. Let the characteristic function of σ be f. The difference in question is then $\sum_{\nu=1}^{n} f(T^{\nu}p) - nm\sigma$. Consider instead the difference $K \cdot \sum_{\nu=1}^{n} f(T^{\nu}p) - n$, where $K = 1/m\sigma$ is the number of cells in the dyadic subdivision to which σ belongs. This latter difference is an integer, and as n increases by unity the difference either increases by $K - 1$ or decreases by 1. To pass from a positive to a negative value it must therefore pass through the value zero. For a random sequence this must happen infinitely often. This means that a random sequence from R has arbitrarily long segments distributed between σ and its complement in *exact* proportion to their measures. We proceed to show that there exist automorphisms of R under which almost all points generate sequences that share this property of random sequences with respect to every dyadic cell, and indeed that such automorphisms are the "general case."

THEOREM 7: *Let R denote the r-dimensional unit cube, $r \geqq 2$. There is a residual set of automorphisms in $M[R]$ under which almost all points p generate sequences distributed in the following manner: Given any dyadic subdivision, there exist arbitrarily large values n such that the first n images of p are equally distributed among the cells of this subdivision.*

Let $\sigma_1^{(j)}, \cdots, \sigma_{N_j}^{(j)}$ be the cells of the j-th dyadic subdivision, and let $f_i^{(j)}$ be the characteristic function of $\sigma_i^{(j)}$. Let $E(i, j, k, n)$ be the set of all T in $M[R]$ such that the measure of the set $A(i, j, n, T)$ of points p for which $\frac{1}{n} \sum_{\nu=1}^{n} f_i^{(j)}(T^{\nu}p) = m\sigma_i^{(j)}$ is greater than $1 - \frac{1}{k}$. The set $E(i, j, k, n)$ is open, because if T belongs to it and G_ϵ denotes the ϵ-neighborhood of the boundary of $\sigma_i^{(j)}$, where ϵ is so chosen that $mG_\epsilon < \frac{1}{n}\left[mA(i, j, n, T) - \left(1 - \frac{1}{k}\right)\right]$, we have $m\left[A(i, j, n, T) - \sum_{\nu=1}^{n} T^{-\nu}G_\epsilon\right] > 1 - \frac{1}{k}$. If p belongs to $A(i, j, n, T) - \sum_{\nu=1}^{n} T^{-\nu}G_\epsilon$, then $Tp, T^2p, \cdots, T^np$ all lie outside G_ϵ, and so $\sum_{\nu=1}^{n} f_i^{(j)}(T_1^{\nu}p) = \sum_{=1}^{n} f_i^{(j)}(T^{\nu}p)$ for any automorphism T_1 so near to T that $\rho(T_1^{\nu}, T^{\nu})$ is less than ϵ for $\nu = 1, \cdots, n$. Consequently, the set $A(i, j, n, T_1)$ contains $A(i, j, n, T) - \sum_{=1}^{n} T^{-\nu}G_\epsilon$, and therefore has measure exceeding $1 - \frac{1}{k}$, so that T_1 belongs to $E(i, j, k, n)$. Hence $\sum_{n=N}^{\infty} \prod_{i=1}^{N_j} E(i, j, k, n)$ is likewise open. But this is the set of automorphisms such

that for some $n \geqq N$ the set of points whose images are distributed equally among $\sigma_1^{(j)}, \cdots, \sigma_{N_j}^{(j)}$ has measure greater than $1 - \frac{1}{k}$. Lemma 18 implies that such automorphisms are everywhere dense in $M[R]$, as may be seen by taking $\alpha > 1 - \frac{1}{k}$, because an automorphism that has a periodic set whose images are equally distributed among $\sigma_1^{(j)}, \cdots, \sigma_{N_j}^{(j)}$ and have combined measure α will belong to $E(i, j, k, n)$ for any value of n divisible by the period of the set. Therefore the set $\prod_{j=1}^{\infty} \prod_{N=1}^{\infty} \prod_{k=1}^{\infty} \sum_{n=N}^{\infty} \prod_{i=1}^{N_j} E(i, j, k, n)$ is residual, since it is an intersection of countably many dense open sets. If T belongs to this set, then, for every pair j, N, almost all points will generate sequences which have segments of length greater than N that are equally distributed among the cells of the j-th dyadic subdivision. The intersection of these sets as j and N run over all positive integers will still have measure one, and the points of this set will have images distributed in the required manner.

13. Automorphisms that preserve sets of measure zero

We shall say that an automorphism *preserves zero sets* if it has the property that $mhA = 0$ if and only if $mA = 0$. This is equivalent to requiring that hA be measurable if and only if A is measurable, because any Lebesgue measurable set is the sum of a Borel set and a set of measure zero, and therefore its image is the sum of a Borel set and the image of a zero set. Consequently, if h preserves zero sets it will also preserve measurable sets. On the other hand, if it does not preserve zero sets either h or its inverse must take some zero set into a set with positive outer measure. Therefore it takes some zero set into a non-measurable set, because every set with positive outer measure contains a non-measurable subset.[43] Hence automorphisms that preserve zero sets might equally well be called *measurability preserving*. It is well-known that an automorphism need not preserve zero sets, but we proceed to show that any automorphism is topologically equivalent to one that does.

THEOREM 8: *Let h be any automorphism of a rectangular r-cell R (or of $E^{(r)}$), $r \geqq 1$. There exists an automorphism g of R (or $E^{(r)}$) such that ghg^{-1} preserves zero sets.*

Consider first the case of an automorphism of R. Define $\mu^*A = \frac{1}{2}m^*A + \sum_{n=1}^{\infty} \frac{1}{2^{n+2}} (m^*h^nA + m^*h^{-n}A)$. Then μ is an r-dimensional LS measure in R, and $\mu R = mR$. By Theorem 2 there exists an automorphism g such that $\mu A = mgA$. Evidently $\mu A = 0$ if and only if $mh^nA = 0$ for every positive or negative integer n. Therefore $\mu hA = 0$ if and only if $\mu A = 0$; that is, $mghg^{-1}A = 0$ if and only if $mA = 0$.

Now consider any automorphism h of $E^{(r)}$. Divide the space into a lattice

[43] Carathéodory, *Vorlesungen über reelle Funktionen*, Leipzig and Berlin 1918, p. 354.

of cubes $R_1, R_2, \cdots$ and let G_i be the interior of R_i. Form the outer measure

$$\mu^*A = \sum_{i=1}^{\infty}\left\{\frac{1}{2}\cdot\frac{m^*AG_i}{mG_i} + \sum_{n=1}^{\infty}\frac{1}{2^{n+2}}\left[\frac{m^*h^n(AG_i)}{mh^nG_i} + \frac{m^*h^{-n}(AG_i)}{mh^{-n}G_i}\right]\right\}.$$

This measure is evidently zero for points and positive for non-void open sets. Furthermore, $\mu R_i = 1$, and therefore bounded sets have finite measure, while $\mu E^{(r)} = +\infty$. Hence, by Theorem 2 Corollary 5, there exists an automorphism g such that $\mu^*A = m^*gA$, at least in case $r \geqq 2$. The same argument as above then shows that ghg^{-1} preserves zero sets. In the case $r = 1$, the positive and negative half lines both have infinite μ-measure, so that the proof is valid in this case also.

14. Sets automorphic to zero sets

We shall say that two subsets A and B of $E^{(r)}$ are automorphic if there exists an automorphism h of $E^{(r)}$ such that $hA = B$. This is evidently a stronger notion than that of homeomorphism of the two sets. We shall obtain a simple characterization of sets that are automorphic to sets of measure zero.

THEOREM 9: *Let B be any subset of $E^{(r)}$, $r \geqq 1$. In order that there exist an automorphism h of $E^{(r)}$ such that hB has Lebesgue measure zero it is necessary and sufficient that the complement of B contain a sequence of perfect sets whose union is dense in $E^{(r)}$. If a bounded set B satisfies this condition, the automorphism h can be taken equal to the identity outside of any cube that contains B.*

The condition is evidently necessary, since if $mhB = 0$ the complement of hB contains a sequence of perfect sets P_n whose union contains all the measure in $E^{(r)}$, and is therefore dense. The perfect sets $h^{-1}P_n$ are therefore outside B and their union is also dense in $E^{(r)}$. To prove that the condition is also sufficient, suppose that B satisfies it and consider any cube R with interior R_0. By hypothesis $R_0 - R_0B$ contains a dense sequence of perfect sets. In $R_0 - R_0B$ it is therefore possible to find a sequence of perfect sets P_n such that every neighborhood in R contains at least one member of the sequence. In each set P_n there exists a Cantor set C_n. Introduce a normalized LS measure μ_n in C_n by mapping it homeomorphically onto a linear Cantor set with linear measure one. Form the outer measure $\mu^*A = mR\cdot\sum_{n=1}^{\infty}\frac{1}{2^n}\mu_n^*AC_n$. This is evidently a LS outer measure in R, zero for points and positive for non-void open sets in R. Also $\mu R = mR$ and $\mu BdR = 0$. By Theorem 2_1 there exists an automorphism $h \in H_0[R]$ such that $\mu^*A = m^*hA$. Since $\mu BR = 0$, we have $mh(BR) = 0$. If B is contained in R we can take h equal to the identity outside R. If B is an unbounded set we can divide $E^{(r)}$ into a lattice of cubes and transform each into itself in such a way as to compress the part of B contained in it to a set of measure zero. The resulting automorphism of $E^{(r)}$ will then carry B into a zero set.

Since any residual set in $E^{(r)}$ contains a dense sequence of perfect sets, it follows as a corollary that any set of first category is automorphic to a zero set.

In a previous paper[44] we have obtained this result directly by a simple category argument. In the case $B \subset R$, we showed in fact that the automorphisms h such that $mhB = 0$ form a residual set in $H[R]$. It is interesting to note that it is *only* for first category sets that this method of proof is available, at least for sets having the property of Baire, because in the same paper it was shown (Theorem 4) that a set having the property of Baire is of first or second category according as the automorphisms that carry it into a set of measure zero form a residual set or a set of first category in $H[R]$. The present theorem may be regarded as completing the earlier result by showing precisely which second category sets are automorphic to zero sets.

15. Transformations topologically equivalent to measure-preserving automorphisms

G. D. Birkhoff has formulated the following problem: Given an automorphism h, when can one assert that there exists an automorphism g such that ghg^{-1} is Lebesgue measure-preserving? In other words, the problem is to characterize *in topological terms* automorphisms that are topologically equivalent to measure-preserving ones. It is well-known that measure-preserving automorphisms have special topological properties, such as the recurrence property discovered by Poincaré and similar topological properties that can be deduced from the ergodic theorem. Again, it is obvious that a measure-preserving automorphism cannot transform a closed sphere into a subset of its interior. Thus it is clear that h must be restricted, in contrast to what we found in the case of the related problem of characterizing automorphisms equivalent to ones that preserve zero sets (§13). In the present section we shall obtain a result that constitutes a solution of this problem, though it must be admitted that the characterization obtained is neither particularly simple nor easy to apply. It is to be hoped that a more elegant solution may be found.

Let us restrict attention to an r-dimensional cube R. If there exists an automorphism g such that ghg^{-1} preserves Lebesgue measure, then h preserves the measure $\mu A = mgA$; and conversely, if h preserves a measure μ automorphic to Lebesgue measure, there will exist a g such that ghg^{-1} preserves Lebesgue measure, namely, any automorphism such that $\mu A = mgA$. Thus Theorem 2 provides the following equivalent formulation of the problem: Given an automorphism h, when can one assert that it preserves some r-dimensional LS measure? It is from this point of view that we shall approach the problem. In a previous paper[45] we have discussed the question of the existence of somewhat more general invariant measures. We obtained the result (loc. cit. Th. 2) that an automorphism T of a complete separable metric space admits a finite invariant Borel measure that is zero for points if (i) T has non-denumerably many

[44] See footnote 28.

[45] J. C. Oxtoby and S. M. Ulam, *On the existence of a measure invariant under a transformation.* Ann. of Math. **40** (1939) 560–566.

periodic points, or (ii) there exists a compact set C consisting of non-periodic points of which at least one returns to C with positive frequency under iteration of T. (A point p of C is said to return with positive frequency if $\lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} f_C(T^\nu p) > 0$, where f_C is the characteristic function of C.) It should be added that in case (i) the invariant measure was defined in such a way that any prescribed sphere containing non-denumerably many periodic points could be made to have positive measure, and in case (ii) positive measure was ascribed to the set C. Assuming this result, we can obtain the following theorem.

THEOREM 10: *Let h be an automorphism of the r-dimensional unit cube R, $r \geqq 2$. In order that there exist an automorphism g such that ghg^{-1} preserves Lebesgue measure it is necessary and sufficient that in every sphere there exist a perfect set to which some point returns with positive frequency under iteration of h, and that these perfect sets can all be chosen outside of an arbitrarily prescribed countable set.*

The necessity of the condition may be deduced from the ergodic theorem as follows. Suppose h is measure-preserving and let A_0 be any given countable set. Then h is a measure-preserving automorphism of the set $E = R - \sum_{n=-\infty}^{+\infty} h^n A_0$, which has measure one. In any sphere there exists a perfect set $P \subset E$ such that $mP > 0$. Denote its characteristic function by $f(p)$. According to the ergodic theorem $f^*(p) = \lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} f(T^\nu p)$ exists for almost all p, and $\int f^*(p)\,dp = mP$. Hence $f^*(p)$ is positive for some point p, that is, p returns to P with positive frequency. Thus h has the required property, and therefore also any equivalent automorphism ghg^{-1}, since the condition is topologically invariant.

Conversely, suppose h fulfills the condition. Let $\sigma_1, \sigma_2, \cdots$ be an enumeration of all dyadic cells in R. Let A_n denote the set of periodic points contained in σ_n provided there are only countably many, otherwise let A_n be void. Then the union A_0 of the sets A_n is countable. Consider any cell σ_n. If it contains non-denumerably many periodic points there exists an invariant measure μ_n which is zero for points and positive for σ_n, in virtue of the theorem quoted above. If σ_n contains only countably many periodic points, then by hypothesis there exists a perfect set $P \subset \sigma_n - A_0$ to which some point returns with positive frequency. Since P consists of non-periodic points, the theorem quoted asserts again the existence of an invariant measure μ_n. Thus we obtain a sequence of invariant Borel measures μ_n, each zero for points and such that $\mu_n\sigma_n > 0$. We suppose them normalized. Then $\mu A = \sum_{n=1}^{\infty} \frac{1}{2^n} \mu_n A$ is a normalized invariant Borel measure in R that is zero for points and positive for non-void open sets. To obtain a measure which in addition is zero for the boundary it is sufficient to form $\nu A = \mu A R_0/\mu R_0$. The Borel measure ν admits an r-dimensional LS extension defined by $\nu^* A = \inf \nu G$, G open, $G \supset A$, which is invariant under h.

By Theorem 2_1 there exists an automorphism g such that $\nu^*A = m^*gA$, and therefore ghg^{-1} preserves Lebesgue measure, as already remarked.

16. A topological ergodic theorem

The paper[45] referred to in the last section also contains the result (Th. 1) that an automorphism T of a complete separable metric space admits a finite invariant Borel measure if there exists a compact set C to which some point returns with positive *superior* frequency $\left(\text{that is, } \limsup_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} f_C(T^\nu p) > 0\right)$. Again it should be added that the measure is positive for C. As a corollary of this, we may assert the following theorem, which properly belongs to the earlier paper since it involves no other results.

THEOREM 11: *Let T be an automorphism of a separable metric space E, and let C be any compact set contained in E. There exists at least one point p in C such that $\lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} f_C(T^\nu p)$ exists, where f_C is the characteristic function of C.*

The interest of this theorem is that, like the ergodic theorem, it asserts the existence of a limiting frequency of return for certain points, but *without any measure-theoretic assumptions.* Nevertheless, the proof requires a very considerable excursion into measure theory. It would be interesting to know whether the result can be obtained directly, even in the special case in which C is a square and T an automorphism of the plane.

To prove the theorem, observe that either $\lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} f_C(T^\nu p) = 0$ for every point of C, in which case the theorem is true, or there exists a point p in C such that $\limsup_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} f_C(T^\nu p) > 0$. In the latter case there exists a finite invariant Borel measure μ such that $\mu C > 0$, by the theorem quoted above. The ergodic theorem is applicable to this measure and asserts that the limit in question exists for almost all points (in this measure) and is positive for at least one point, since its integral is equal to μC. Therefore there exists a point of C, not necessarily equal to p, which returns with positive frequency.

17. Connections with Haar measure

Haar's Theorem[46] asserts that in any locally compact separable topological group there exists a left (or right) invariant measure which is finite and positive for every compact neighborhood. It is evidently zero for points, and it is unique up to a constant multiplier.[47] In the compact case it is finite, and then left and

[46] A. Haar, *Der Massbegriff in der Theorie der kontinuierlichen Gruppen.* Ann. of Math. **34** (1933) 147–169.

[47] J. von Neumann, *The uniqueness of Haar's measure.* Recueil Mathématique, Moscou, N.S. **1** (1936) 721–734.

right invariant measures coincide and are also inverse invariant.[48] The measure is derived from a Carathéodory outer measure that satisfies condition M5.[49] In case the group manifold is a polyhedron, the Haar measure is therefore r-dimensional, and Theorem 2 Corollary 1 establishes the following connection with Lebesgue measure. *If the group manifold of a topological group is a finite euclidean polyhedron, the Haar measure, suitably normalized, is automorphic to the Lebesgue measure in the polyhedron.* Again, we may say that *if two groups have for group manifold the same polyhedron, their Haar measures are automorphic to each other, provided only that they are similarly normalized.*

18. Approximation by locally linear automorphisms

In the proof of Theorem 1, methods were developed that made it possible to approximate any measure-preserving automorphism by one that has a periodic Cantor set whose images are distributed in a regular manner. It is natural to ask whether one can find similarly an approximating automorphism that has a periodic cube. In the present section we shall show that this is possible for automorphisms of a cube that leave the boundary fixed, indeed that an approximating automorphism can be found that has a sequence of periodic cubes which together include almost all points and which the transformation permutes among themselves as if by translation.

It may be remarked that the same methods can be extended to apply to r-dimensional polyhedra embedded in r-space, and to automorphisms that are merely isotopic to the identity. An attempt to extend them to general automorphisms and polyhedra seems to encounter more serious difficulties, and for this reason the proof of Theorem 1 was based on periodic Cantor sets rather than on periodic cubes, which might have seemed more natural.

LEMMA 19: *Let R be the r-dimensional unit cube, $r \geqq 2$, and let T be a measure-preserving automorphism of R that has an invariant set consisting of a finite number of disjoint cubes $R_1, \cdots, R_M$ interior to R and each a sum of cubes of some dyadic subdivision. There exists an arbitrarily small automorphism $S \in M_0[R]$, equal to the identity on $R_1, \cdots, R_M$, such that ST permutes cyclically the centers of the cubes of a certain arbitrarily fine dyadic subdivision of R that are not contained in $R_1, \cdots, R_M$.*

Let E be the polyhedron obtained from R by removing the interiors of $R_1, \cdots, R_M$. Then E is regularly connected, and T may be considered as an automorphism of E. By the Corollary to Theorem 1, T can be made metrically transitive by composing it with an arbitrarily small automorphism that leaves the boundary of E fixed. We may therefore assume without loss of generality that T itself is metrically transitive with respect to E. By Lemma 3, we can represent E as the continuous image of a rectangular r-cell R^* under a measure-

[48] J. von Neumann, *Zum Haarschen Mass in topologischen Gruppen.* Compositio Mathematica **1** (1934) 106–114.

[49] See the note of S. Banach in Saks, *Theory of the Integral*, Appendix 1.

preserving map f which is a homeomorphism up to the boundary. Let ϵ be a given positive number and let $\delta > 0$ be such that any automorphism in $H_0[R^*]$ with norm less than δ corresponds to an automorphism of E with norm less than ϵ. Let $\sigma_1, \cdots, \sigma_N$ be the cubes in E belonging to a dyadic subdivision of diameter less than ϵ. Order these in a sequence $\sigma_{i_1}, \cdots, \sigma_{i_K}$ such that each cube σ_i appears at least once and such that σ_{i_k} and $\sigma_{i_{k+1}}$ have a regular face in common, $1 \leqq k \leqq K - 1$. (It may be possible to find such an ordering without repetitions, but this is not necessary.) Since T is metrically transitive on E there exists a non-periodic point whose images under iteration of T fall in the interiors of each of the cubes σ_i with frequency $1/N$, in fact, such points form a set of measure mE. Let p be such a point, then for any sufficiently large n the number of points $Tp, \cdots, T^n p$ in any cube σ_i will differ from n/N by an arbitrarily small fraction of n, in particular by less than n/NK^2. In addition, let us choose n so that $n > NK^2\lambda$ and so that $n + \lambda$ is a number of the form $N \cdot 2^{\alpha r}$, where λ is the integer corresponding to δ defined in Lemma 14. As in the proof of Lemma 15, we can modify the transformation $f^{-1}Tf$ by composition with an automorphism $S_0 \in M_0[R^*]$ such that under $S_0 f^{-1}Tf$ the point $f^{-1}p$ has period $n' = n + \lambda$ and its first n images are the same as under $f^{-1}Tf$. (Let D be the set on which $T_0 = f^{-1}Tf$ is 1:1, and let $p_i^* = f^{-1}T^i p$, $0 \leqq i \leqq n$. Then $p_n^* \in (T_0^{-1}U_\delta)p_0^*$ and step by step we can choose points $p_{n+1}^*, \cdots, p_{n+\lambda-1}^*$ such that $p_0^*, \cdots, p_{n+\lambda-1}^*$ are distinct, and therefore also $T_0 p_0^*, \cdots, T_0 p_{n+\lambda-1}^*$, and such that $T_0 p_i^*$ is always within distance δ of p_{i+1}^*, $p_{n+\lambda}^*$ being taken equal to p_0^*. Then define S_0 with norm less than δ so that $S_0 T_0 p_i^* = p_{i+1}^*$, $0 \leqq i \leqq n + \lambda - 1$.) Then $S_1 = fS_0 f^{-1}$ will be an automorphism of E with norm less than ϵ that leaves the boundary fixed, and under $S_1 T$ the point p will have period n' and its first n images will still be $Tp, \cdots, T^n p$. It follows that the number of distinct images of p under $S_1 T$ in any cube σ_i differs from n'/N by less than $\eta = 2n'/NK^2$, because if σ_i contains n_i of the points $Tp, \cdots, T^n p$, and λ_i of the points $T_1^{n+1}p, \cdots, T_1^{n+\lambda}p$, we have

$$\left| n_i + \lambda_i - \frac{n'}{N} \right| \leqq \left| n_i - \frac{n}{N} \right| + \left| \lambda_i - \frac{\lambda}{N} \right| \leqq \frac{n}{NK^2} + \lambda < \frac{2n}{NK^2} < \eta.$$

Let $p_1, \cdots, p_{n'}$ denote the n' distinct images of p under $S_1 T$. We proceed to assign each of these to one of the cubes σ_i in such a way that exactly n'/N are assigned to each cube σ_i, and each point is assigned either to the cube in which it lies or to an adjacent one. To do this we proceed as follows. First assign each point tentatively to the cube in which it lies. If the number of points in σ_{i_1} is less than n'/N, make up the deficiency by assigning to it points from σ_{i_2}. If the number in σ_{i_1} is greater than n'/N, assign enough points from it to σ_{i_2} to remove the excess. Next consider σ_{i_2} and eliminate the excess or deficiency of points now assigned to it by assigning points of it to σ_{i_3}, or to it from σ_{i_3}. After the k-th step there will be exactly n'/N points assigned to each of the cubes $\sigma_{i_1}, \sigma_{i_2}, \cdots, \sigma_{i_k}$ except those (if any) which are equal to $\sigma_{i_{k+1}}$. After the $(K - 1)$-st step there will be exactly n'/N points assigned to each cell

σ_i. At the first step the number of points assigned to other cubes is at most η, at the second step at most 2η are assigned, and at the last step at most $(K - 1)\eta$. The total number of points assigned to other cubes is therefore at most $\frac{1}{2}K(K - 1)\eta$, which is less than the number of points p_j in any one of the cubes σ_i. Hence the assignments can be made in such a way that no point is assigned to another cube more than once, and therefore every point is assigned finally either to the cube in which it lies or to an adjacent one. We can therefore number the centers $q_1, \cdots, q_{n'}$ of the cubes of the α-th dyadic subdivision of $\sigma_1, \cdots, \sigma_N$ in such a way that the points q_j and p_j always lie in the same cube σ_i or in cubes having an $(r - 1)$-face in common. The distance between them is therefore less than 2ϵ, and we can define a measure-preserving automorphism S_2 of E that leaves the boundary fixed, has norm less than 2ϵ, and carries each p_j into q_j. (This does not follow directly from Lemma 13 as stated, but the same method of proof used there evidently applies in the present case since the line segment joining p_j to q_j lies in the interior of E.) It follows that the points $q_1, \cdots, q_{n'}$ are permuted cyclically by $S_2S_1TS_2^{-1}$, which may be written in the form ST by letting $S = S_2S_1TS_2^{-1}T^{-1}$. This completes the proof, since S leaves the entire boundary of E fixed and can be made arbitrarily small by suitable choice of ϵ, since S_1 and S_2 have norms less than 2ϵ.

LEMMA 20: *Let R be the r-dimensional unit cube, $r \geqq 2$, and let $T \in M_0[R]$ be such that it permutes cyclically the centers of some of the cubes of a certain dyadic subdivision of R. Let $\sigma_1, \cdots, \sigma_N$ be the cubes whose centers are permuted. There exist automorphisms $S \in M_0[R]$ and $h \in H_0[R]$, equal to the identity outside the cubes σ_i and on their boundaries, such that $hSTh^{-1}$ is measure-preserving and under it the cubes concentric*[50] *to $\sigma_1, \cdots, \sigma_N$ with half their diameter are rigidly permuted in a single cycle.*

Let p_i be the center of σ_i. We may suppose the cubes so numbered that $Tp_i = p_{i+1}$, $i < N$, and $Tp_N = p_1$. Let $\tau_1, \cdots, \tau_N$ be cubes concentric to $\sigma_1, \cdots, \sigma_N$, all with the same diameter, and small enough so that $T\tau_i$ is interior to σ_{i+1}, $i < N$, and $T\tau_N$ to σ_1. Let g_1 be an automorphism of R that leaves the boundary fixed and translates τ_1 into τ_2. Extend both g_1 and T to the whole space $E^{(r)}$ by defining them equal to the identity outside R. Let g_2 be a radial shrinking transformation of $E^{(r)}$ such that $g_2R \subset \tau_2$ and such that g_2 leaves both τ_2 and $T\tau_1$ fixed. It is easily verified that $(g_2g_1T^{-1}g_2^{-1})T$ translates τ_1 into τ_2, and $g_2g_1T^{-1}g_2^{-1}$ is an automorphism equal to the identity outside τ_2 that carries $T\tau_1$ into τ_2 in a measure-preserving manner. By Theorem 2 Corollary 3, we can modify $g_2g_1T^{-1}g_2^{-1}$ in $\tau_2 - T\tau_1$ in such a way that it becomes a measure-preserving automorphism of τ_2 that leaves the boundary fixed. Call this modified automorphism S_2 and define it equal to the identity outside τ_2. Then S_2T carries τ_1 rigidly into τ_2. In exactly the same way we can define $S_3 \in M_0[\tau_3]$ such that S_3T carries τ_2 rigidly into τ_3. Finally we define $S_1 \in M_0[\tau_1]$

[50] We shall use the term "concentric" to mean that the cubes have the same center and also have corresponding faces parallel.

such that S_1T carries τ_N rigidly into τ_1. The product of these N transformations S_i defines an automorphism S of R that carries each cube σ_i into itself and is equal to the identity on their boundaries and outside them. Under ST each cube τ_i is translated into τ_{i+1}, $i < N$, and τ_N into τ_1. Finally, let h be a radial transformation of each cube σ_i into itself which carries τ_i into the concentric cube σ_i^* with diameter half that of σ_i. Then $hSTh^{-1}$ is a measure-preserving automorphism that permutes the cubes σ_i^* rigidly and cyclically.

LEMMA 21: *Let T be a measure-preserving automorphism of the r-dimensional unit cube R, $r \geqq 2$, and suppose that T permutes rigidly certain disjoint cubes $R_1, \cdots, R_K$ contained in the interior of R which are sums of cubes of a dyadic subdivision. There exists an arbitrarily small automorphism $S \in M_0[R]$ equal to the identity on $R_1, \cdots, R_K$ such that ST permutes rigidly in a single cycle the cubes concentric to and with half the diameter of those of a certain dyadic subdivision of the complementary set $E = R - (R_1 + \cdots + R_K)$.*

We shall consider the degenerate case $K = 0$ to be included in this statement.

Let $\epsilon > 0$ be given. By the continuity of the group product, there exists a positive number δ such that if h, S_1, S_2 are any automorphisms with norm less than δ then $hS_2S_1Th^{-1}T^{-1}$ will have norm less than ϵ. By Lemma 19 (or by Lemma 15 in the case $K = 0$), there exists $S_1 \in M_0[R]$ equal to the identity on $R_1, \cdots, R_K$ with norm less than δ such that S_1T permutes cyclically the centers of the cubes $\sigma_1, \cdots, \sigma_N$ of a dyadic subdivision of E, and we may suppose the cubes σ_i to have diameter less than δ. By Lemma 20, there exist automorphisms h and S_2 that carry these cubes into themselves and leave their boundaries fixed such that $hS_2(S_1T)h^{-1}$ is measure-preserving and permutes rigidly in a single cycle the cubes σ_i^* concentric to and with half the diameter of σ_i. But this transformation may be written in the form ST, where $S = hS_2S_1Th^{-1}T^{-1}$. The automorphism S is measure-preserving, since it is a product of two such transformations; it is equal to the identity on $R_1, \cdots, R_K$ and on the boundary of R, since h, S_1, S_2 are; and it has norm less than ϵ, since h, S_1, S_2 have norms less than δ.

THEOREM 12: *Any measure-preserving automorphism T of the r-dimensional unit cube R, $r \geqq 2$, that leaves the boundary fixed can be approximated uniformly by another such automorphism T^* which is locally linear almost everywhere. More precisely, there is a sequence of disjoint cubes interior to R, having total measure one, which under T^* are rigidly permuted among themselves.*

The proof is based on a sequence of modifications of T defined inductively. Let $\delta_1, \delta_2, \cdots$ be a sequence of positive numbers with sum less than a given number $\epsilon > 0$. We shall show that it is possible to find a sequence of automorphisms $S_1, S_2, \cdots$ with the following properties:

1° $S_n \in M_0[R]$, $\rho(S_n, I) < \delta_n$

2° $S_nS_{n-1} \cdots S_1T$ permutes rigidly among themselves a finite number of disjoint cubes $R_1, \cdots, R_{N_n}$ in n cycles, each of these cubes being a sum of interior cubes of a certain dyadic subdivision of R.

3° S_n is equal to the identity on $R_1 + R_2 + \cdots + R_{N_{n-1}}$, $n > 1$.

4° $m(R_1 + \cdots + R_{N_n}) = 1 - \left(1 - \frac{1}{2^r}\right)^n$.

By Lemma 21, there exists $S_1 \epsilon M_0[R]$ with norm less than δ_1 such that S_1T permutes rigidly in a single cycle the cubes $R_1, \cdots, R_{N_1}$ concentric to and with half the diameter of those of a certain dyadic subdivision. These cubes therefore have total measure $1/2^r$. Thus conditions 1° to 4° are satisfied in the case $n = 1$, condition 3° being vacuous. Now suppose that $S_1, \cdots, S_n$ and $R_1, \cdots, R_{N_n}$ have been defined so that conditions 1° to 4° are satisfied for values of $n \leqq k$. Then $T_k = S_kS_{k-1} \cdots S_1T$ fulfills the hypotheses of Lemma 21 and there exists an automorphism $S_{k+1} \epsilon M_0[R]$, with norm less than δ_{k+1}, equal to the identity on $R_1 + \cdots + R_{N_k}$, such that $S_{k+1}T_k$ permutes rigidly in a single cycle the cubes $R_{N_k+1}, \cdots, R_{N_{k+1}}$ concentric to and with half the diameter of those of a certain dyadic subdivision of R that are not contained in $R_1 + \cdots + R_{N_k}$. The total measure of these additional cubes is therefore $\frac{1}{2^r}\left(1 - \frac{1}{2^r}\right)$, so that conditions 1° to 4° are now satisfied for values of $n \leqq k + 1$, and the inductive definition of the successive modifications S_n is complete.

From 1° it follows that the limit $S = \lim_{n\to\infty} S_nS_{n-1} \cdots S_1$ exists and has norm less than ϵ. Furthermore, ST is equal to $S_n \cdots S_1T$ on $R_1 + R_2 + \cdots + R_{N_n}$, in virtue of 3°, and so it permutes these cubes rigidly among themselves. Thus under ST the cubes $R_1, R_2, \cdots$ are all rigidly permuted among themselves in finite cycles. Finally, by 4°, these cubes contain almost all points of R. Hence $T^* = ST$ has all the properties described in the theorem.

Bryn Mawr College
The University of Wisconsin

Annals of Mathematics
Vol. 42, No. 4, October, 1941

A REMARK ON NORMAL VARIETIES

By H. T. Muhly*

(Received December 22, 1940)

1. In the terminology of the Italian School an algebraic variety is called "normal" if its system of hyperplane sections is complete. O. Zariski applies the term "normal" to an algebraic variety whose associated ring of homogeneous coordinates is integrally closed.[1] The two concepts are not equivalent. Zariski refers to a variety which satisfies the former condition as "normal in the geometric sense" and to one which satisfies the latter condition as "normal in the arithmetic sense."

A variety which is normal in the arithmetic sense is necessarily normal in the geometric sense. Moreover, an r-dimensional variety V_r which is normal in the arithmetic sense has no $(r - 1)$-dimensional singularities. A variety which is normal in the geometric sense need not be normal in the arithmetic sense. For example, a plane quartic of genus two is normal in the geometric sense, but has a double point. A curve may be free from singularities and yet not normal in the arithmetic sense. A rational space quartic illustrates this possibility. The validity of all of these assertions is established in Z.

The object of this note is to characterize geometrically those algebraic varieties which are normal in the arithmetic sense. To this end we propose the following theorem: *A necessary and sufficient condition that the r-dimensional algebraic variety V_r be normal in its ambient projective space P_n is that for every integer m the linear system cut out on V_r by the hypersurfaces of order m in P_n be complete.*

In the course of the proof we need the notion of the "character of homogeneity" of an algebraic variety, introduced by Zariski in the paper Z. Let $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ be the homogeneous coordinates of the general point of a variety V_r in the projective space P_n. The underlying field of constants for P_n is assumed to be algebraically closed and of characteristic zero. We denote this field by K. The integral closure, $K[\zeta_0^*, \zeta_1^*, \cdots, \zeta_h^*]$ of the ring of homogeneous coordinates $K[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$ in its quotient field Σ^* is denoted by $\bar{\mathfrak{o}}^*$. Each of the quantities $\zeta_0^*, \zeta_1^*, \cdots, \zeta_h^*$ may be assumed to be homogeneous of positive degree.[2] Zariski proves that there exist integers δ with the following property.

* National Research Fellow, 1940–41.

[1] O. Zariski, "Some Results in the Arithmetic Theory of Algebraic Varieties," American Journal of Mathematics, Vol. LXI, No. 2 (April 1939); designated henceforth by Z.

[2] If $\xi_i = \xi_i^*/\xi_0^*$, then $\xi_1, \xi_2, \cdots, \xi_n$ are the non-homogeneous coordinates of the general point of our V_r and $\Sigma(=K(\xi_1, \cdots, \xi_n))$ is the field of rational functions on V_r. The field Σ^* is then equal to $\Sigma(\xi_0)$, a pure transcendental extension of Σ. An element ω^* of Σ^* is said to be homogeneous of degree ν if $\tau \times \omega^* = t^\nu \omega^*$ for every automorphism τ of Σ^* over Σ of the form $\tau : \xi_0^* \to t\xi_0^*, t \in K$.

If $\omega_0^*, \omega_1^*, \cdots, \omega_m^*$ denote the possible power products of $\zeta_0^*, \zeta_1^*, \cdots, \zeta_h^*$ which are homogeneous of degree δ and if ω^* is any power product of $\zeta_0^*, \zeta_1^*, \cdots, \zeta_h^*$ of degree of homogeneity $\rho\delta$ (ρ a positive integer) then ω^* is a form of degree ρ in $\omega_0^*, \omega_1^*, \cdots, \omega_m^*$. Any integer δ with this property is called a character of homogeneity of V_r.

2. The necessity of our condition is easily established. The ring of homogeneous coordinates $\mathfrak{o}^* = K[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$ is integrally closed in its quotient field Σ^* since we assume V_r to be normal in the arithmetic sense. As we have already pointed out, our hypothesis assures us that the linear system $| V |$ which is cut out on V_r by the hyperplanes of P_n is complete. To prove that all of the multiples, $| mV |$, $m = 1, 2, \cdots$, of $| V |$ are also complete we point out that the integer one is a character of homogeneity of V_r since $\mathfrak{o}^*$ is itself integrally closed and since ξ_i^* is homogeneous of degree 1, $i = 0, 1, \cdots, n$. It follows by sections 20 and 21 of Z that if $\omega_0^*, \omega_1^*, \cdots, \omega_h^*$ are the linearly independent power products of degree m (m a positive integer) which occur among all the power products of degree m which can be formed from $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$, then $\omega_0^*, \omega_1^*, \cdots, \omega_h^*$ are the homogeneous coordinates of the general point of a variety V_r' which is normal in the arithmetic sense and which is birationally equivalent to V_r. Moreover, in the birational correspondence between V_r and V_r' the system of sections of V_r with the hypersurfaces of order m is transformed into the system of sections of V_r' with the hyperplanes of its ambient space P_h. Since V_r' is normal in the arithmetic sense the system of hyperplane sections of V_r' is complete. This proves the first half of our theorem.

3. Before proving that our condition is sufficient we list certain well known definitions and results to which we shall have occasion to refer. a) If $\Sigma = K(\xi_1, \cdots, \xi_n)$ is the field of rational functions on V_r, then by a prime divisor of Σ we shall mean an $(r - 1)$-dimensional valuation of Σ. If $\mathfrak{J}$ is any finite integral domain in Σ, and if $\mathfrak{P}$ is a prime divisor whose valuation ring $\mathfrak{B}_\mathfrak{P}$ contains $\mathfrak{J}$, then $\mathfrak{P}$ will be said to be of the first kind with respect to $\mathfrak{J}$ if the origin (the prime ideal of elements of positive value in $\mathfrak{J}$) of $\mathfrak{P}$ in $\mathfrak{J}$ is an $(r - 1)$-dimensional ideal. If this is not the case then $\mathfrak{P}$ is of the second kind with respect to $\mathfrak{J}$.

b) Let $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ be the homogeneous coordinates of the general point of a variety V_r in P_n, and let $\mathfrak{o}^* = K[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$ be the ring associated with $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$. We may assume that $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$ are algebraically independent and that $\xi_{r+1}^*, \cdots, \xi_n^*$ depend integrally on $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$. This situation may always be realized by means of a non-singular projective transformation. If we put $\xi_i = \xi_i^*/\xi_0^*$, $\mathfrak{o}_0 = K[\xi_1, \xi_2, \cdots, \xi_n]$, $\mathfrak{o}_i = K[\xi_1/\xi_i, \cdots, \xi_{i-1}/\xi_i, 1/\xi_i, \xi_{i+1}/\xi_i, \cdots, \xi_n/\xi_i]$, then it is not difficult to see that in each of the rings $\mathfrak{o}_j$, $j = 0, 1, \cdots, r$, the first r coordinates are algebraically independent and that the remaining ones depend integrally on the first r (see section 4). The valuation ring $\mathfrak{B}_\mathfrak{P}$ of any prime divisor, $\mathfrak{P}$, must contain at least one of the rings $\mathfrak{o}_j$, $j = 0, 1, \cdots, r$. Moreover, it is easily seen that if $\mathfrak{o}_\alpha \subset \mathfrak{B}_\mathfrak{P}$ and

$\mathfrak{o}_\beta \subset \mathfrak{B}_\mathfrak{P}$, and if $\mathfrak{P}$ is of the first kind with respect to $\mathfrak{o}_\alpha$ then it is also of the first kind with respect to $\mathfrak{o}_\beta$. Hence we say that $\mathfrak{P}$ is of the first kind (or of the second kind) with respect to the given projective model V_r of Σ.

We consider the set $\mathfrak{S}$ of all prime divisors of Σ which are of the first kind with respect to V_r, and we form the abelian group $\mathfrak{G}(V_r)$ which consists of all finite formal power products $\mathfrak{P}_1^{\alpha_1} \cdot \mathfrak{P}_2^{\alpha_2} \cdots \mathfrak{P}_k^{\alpha_k}$, $\mathfrak{P}_i \in \mathfrak{S}$, α_i an integer, $i = 1, 2, \cdots, k$. The elements of $\mathfrak{G}(V_r)$ are called divisors. The divisor $\mathfrak{A} = \mathfrak{P}_1^{\alpha_1} \cdot \mathfrak{P}_2^{\alpha_2} \cdot \cdots \cdot \mathfrak{P}_k^{\alpha_k}$ is said to be integral if $\alpha_i > 0$, $i = 1, 2, \cdots, k$.

As in the case of functions of one variable, a uniquely determined divisor $\mathfrak{A}(\eta)$ may be associated with each element η in Σ. We have

$$\mathfrak{A}(\eta) = \prod \mathfrak{P}^{v_\mathfrak{P}(\eta)}$$

where the product is extended over all prime divisors in $\mathfrak{S}$. We point out that, since every $\mathfrak{P}$ in $\mathfrak{S}$ is an $(r - 1)$-dimensional valuation of Σ, the value group of $\mathfrak{P}$ may be taken to be the group of integers. Those divisors $\mathfrak{A}$ in $\mathfrak{G}(V_r)$ which are associated in this way with elements η in Σ form a subgroup $\mathfrak{H}$ of $\mathfrak{G}(V_r)$. Two divisors $\mathfrak{A}$ and $\mathfrak{B}$ are said to be equivalent if $\mathfrak{A} \cdot \mathfrak{B}^{-1} \in \mathfrak{H}$.

With every prime divisor $\mathfrak{P}$ in $\mathfrak{S}$, one can associate a unique irreducible $(r - 1)$-dimensional subvariety $V(\mathfrak{P})$ of V_r. With every integral divisor $\mathfrak{A} = \mathfrak{P}_1^{\alpha_1} \cdot \cdots \cdot \mathfrak{P}_k^{\alpha_k}$, one associates the effective $(r - 1)$-dimensional subvariety

$$V(\mathfrak{A}) = \alpha_1 V(\mathfrak{P}_1) + \cdots + \alpha_k V(\mathfrak{P}_k).$$

One then defines the complete system determined by $V(\mathfrak{A})$ to be the totality of all effective curves associated with integral divisors which are equivalent to $\mathfrak{A}$. Zariski's definition of a complete system, which we have used until now, is equivalent to this one in so far as the system of hyperplane sections and its multiples on a given V_r is concerned (see Z page 288).

c) LEMMA: *The algebraically independent quantities $\xi_1, \xi_2, \cdots, \xi_r$ have divisor representations of the form*

$$\xi_i = \mathfrak{A}_i/\mathfrak{A}, \qquad i = 1, 2, \cdots, r,$$

where $\mathfrak{A}_i$ and $\mathfrak{A}$ are integral divisors and no prime divisor divides both $\mathfrak{A}_i$ and $\mathfrak{A}$. The proof of the lemma runs along the following lines. Let $\mathfrak{P}$ be a prime divisor, of the set $\mathfrak{S}$, which is such that for at least one i the inequality $v_\mathfrak{P}(\xi_i) < 0$ holds. We may assume that $v_\mathfrak{P}(\xi_i) \leqq v_\mathfrak{P}(\xi_j)$, $j = 1, 2, \cdots, r$, for if this is not the case we consider the ξ of least value. The valuation ring $\mathfrak{B}_\mathfrak{P}$ must then contain the ring $\mathfrak{o}_i = K[\xi_1/\xi_i, \cdots, \xi_{i-1}/\xi_i, 1/\xi_i, \xi_{i+1}/\xi_i, \cdots, \xi_n/\xi_i]$. Moreover, since $v_\mathfrak{P}(1/\xi_i) > 0$, $1/\xi_i \equiv 0(\mathfrak{p})$, where $\mathfrak{p}$ is the origin of $\mathfrak{P}$ in $\mathfrak{o}_i$. Since $\mathfrak{p}$ is $(r - 1)$-dimensional, $\mathfrak{p}$ cannot divide ξ_j/ξ_i, $j = 1, 2, \cdots, r$. (This follows since $\xi_{r+1}/\xi_i, \cdots, \xi_n/\xi_i$ depend integrally on $\xi_1/\xi_i, \cdots, \xi_{i-1}/\xi_i, 1/\xi_i, \xi_{i+1}/\xi_i, \cdots, \xi_r/\xi_i$.) It therefore follows that $v_\mathfrak{P}(\xi_j/\xi_i) = 0$, or that $v_\mathfrak{P}(\xi_j) = v_\mathfrak{P}(\xi_i)$, $j = 1, 2, \cdots, r$. This proves the lemma.

4. We prove that our condition is sufficient. Let V_r be a variety in the projective space P_n, and let $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ be the homogeneous coordinates

of its general point. We assume that the system $|mV|$ cut out on V_r by the hypersurfaces of order m in P_n is complete, $m = 1, 2, \cdots$. We must prove that the ring $\mathfrak{o}^* = K[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$ is integrally closed in its quotient field Σ^*. As in section 3, we assume that $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$ are algebraically independent and that $\xi_{r+1}^*, \cdots, \xi_n^*$ depend integrally on $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$. Any element of Σ^* which depends integrally on $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$ is the sum of homogeneous elements which depend integrally on $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$. It is therefore sufficient to prove that $\mathfrak{o}^*$ contains any homogeneous element of Σ^* which depends integrally on $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$.

Let ω^* be an homogeneous element of degree ν with this property and let

$$\omega^{*h} + a_1(\xi_0^*, \xi_1^*, \cdots, \xi_r^*)\omega^{*h-1} + \cdots + a_h(\xi_0^*, \xi_1^*, \cdots, \xi_r^*) = 0 \tag{1}$$

be the equation which expresses the integral dependence of ω^* on $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$. Since every homogeneous integer satisfies an homogeneous equation of integral dependence, we may assume that a_i is a form of degree $i\nu$ in $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$, $i = 1, 2, \cdots, h$ (see Z, section 18). If we divide equation (1) by $\xi_0^{*h\nu}$ we see that the quantity $\omega = \omega^*/\xi_0^{*\nu}$ is an element of Σ $(= K(\xi_1, \cdots, \xi_n))$ which depends integrally on $\xi_1, \xi_2, \cdots, \xi_r$. On the other hand, division of equation (1) by $\xi_i^{*h\nu}$ shows that ω/ξ_i^ν depends integrally on $\xi_1/\xi_i, \cdots, \xi_{i-1}/\xi_i, 1/\xi_i, \xi_{i+1}/\xi_i, \cdots, \xi_r/\xi_i$, $i = 1, 2, \cdots, r$.

Let $\xi_i = \mathfrak{A}_i/\mathfrak{A}$, $i = 1, 2, \cdots, r$, be the divisor decomposition of ξ_i mentioned in 3c. Since the system $|\nu V|$ is complete for every ν, it follows that if $\mathfrak{B}$ is an integral divisor in $\mathfrak{G}(V_r)$ which is equivalent to $\mathfrak{A}^\nu$ (in symbols $\mathfrak{B} \sim \mathfrak{A}^\nu$), and if $\theta = \mathfrak{B}/\mathfrak{A}^\nu$, then $\theta = f(\xi_1, \cdots, \xi_n)$ where f is a polynomial of degree $\leqq \nu$ in the indicated arguments.

If $\theta = \mathfrak{B}/\mathfrak{A}^\nu$, then θ is an integral function of $\xi_1, \xi_2, \cdots, \xi_r$, since θ has non-negative value at all prime divisors of $\mathfrak{S}$ for which $\xi_1, \xi_2, \cdots, \xi_r$ have non-negative values. On the other hand, $\theta/\xi_i^\nu = \mathfrak{B}/\mathfrak{A}_i^\nu$, and consequently θ/ξ_i^ν is an integral function of $\xi_1/\xi_i, \cdots, 1/\xi_i, \cdots, \xi_r/\xi_i$. We assert that, conversely, if an integral function θ of $\xi_1, \cdots, \xi_r$ has the property that for some integer ν, θ/ξ_i^ν is an integral function of $\xi_1/\xi_i, \cdots, \xi_{i-1}/\xi_i, 1/\xi_i, \xi_{i+1}/\xi_i, \cdots, \xi_r/\xi_i$, $i = 1, 2, \cdots, r$, then θ may be put into the form $\mathfrak{B}/\mathfrak{A}^\nu$ where $\mathfrak{B}$ is an integral divisor in $\mathfrak{G}(V_r)$ which is necessarily equivalent to $\mathfrak{A}^\nu$. In fact, if $\theta = \mathfrak{B}_0/\mathfrak{A}_0$ is the divisor decomposition of θ, then $\theta/\xi_i^\nu = \mathfrak{B}_0\mathfrak{A}^\nu/\mathfrak{A}_0\mathfrak{A}_i^\nu$. Since θ is an integral function of $\xi_1, \cdots, \xi_r$, no prime divisor of the set $\mathfrak{S}$ can divide $\mathfrak{A}_0$ if it does not divide $\mathfrak{A}$. Moreover, no prime divisor of $\mathfrak{S}$ can occur in $\mathfrak{A}_0$ to a higher power than that to which it occurs in $\mathfrak{A}^\nu$, since θ/ξ_i^ν is an integral function of $\xi_1/\xi_i, \cdots, \xi_{i-1}/\xi_i, 1/\xi_i, \xi_{i+1}/\xi_i, \cdots, \xi_r/\xi_i$, $i = 1, 2, \cdots, r$. It follows that $\mathfrak{A}^\nu = \mathfrak{L} \cdot \mathfrak{A}_0$ where $\mathfrak{L}$ is an integral divisor. We can therefore write $\theta = \mathfrak{B}/\mathfrak{A}^\nu$.

Our theorem now follows readily. If ω^* is an element of Σ^* which depends integrally on $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$ and if ω^* is homogeneous of degree ν, then the quantity $\omega = \omega^*/\xi_0^{*\nu}$ is an integral function of $\xi_1, \xi_2, \cdots, \xi_r$ which has the property that ω/ξ_i^ν is an integral function of $\xi_1/\xi_i, \cdots, \xi_{i-1}/\xi_i, 1/\xi_i, \xi_{i+1}/\xi_i, \cdots, \xi_r/\xi_i$. The integral function ω therefore admits a divisor representation of the form $\mathfrak{B}/\mathfrak{A}^\nu$.

By virtue of the completeness of $| \nu V |$, this implies that $\omega = f(\xi_1, \cdots, \xi_n)$, where f is a polynomial of degree $\leqq \nu$. Hence we see that

$$\omega^* = \xi_0^{*\nu} \cdot f(\xi_1, \cdots, \xi_n) = \varphi_\nu(\xi_0^*, \xi_1^*, \cdots, \xi_n^*)$$

where φ_ν is a form of degree ν. The element ω^* is thus seen to be in $\mathfrak{o}^*$, q.e.d.

As a final remark we point out that there exist curves which are free from singularities and on which the hyperplanes cut out a complete system, but which are nevertheless not normal in the arithmetic sense. In fact in P_3 there exist curves of order seven and of genus four which are free from singularities. On such curves the system of plane sections is complete while the system cut out by the quadric surfaces is not. Thus, even though freedom from singularities and completeness of the system of plane sections are individually necessary for a curve to be normal in the arithmetic sense, they are together not sufficient.

PRINCETON UNIVERSITY AND INSTITUTE FOR ADVANCED STUDY

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

ON THE CONNECTION BETWEEN THE ORDINARY AND THE MODULAR CHARACTERS OF GROUPS OF FINITE ORDER*

BY RICHARD BRAUER

(Received October 21, 1940)

Introduction

The representations of groups by matrices with coefficients in a modular field and the corresponding modular group characters have been studied in two earlier papers;[1] the aim of the present paper is to continue this work. Let $\mathfrak{G}$ be a group of finite order g and let p be a rational prime. If $\zeta_1(G)$, $\zeta_2(G)$, $\cdots$ are the (ordinary) characters of $\mathfrak{G}$, and $\varphi_1(G)$, $\varphi_2(G)$, $\cdots$ the modular characters of $\mathfrak{G}$ for p, then we have formulae

$$(*) \qquad \zeta_\mu(G) = \sum_\nu d_{\mu\nu}\varphi_\nu(G)$$

provided that G is a p-regular element of $\mathfrak{G}$, i.e. an element G of an order prime to p. The $d_{\mu\nu}$ are non-negative rational integers, the decomposition numbers of $\mathfrak{G}$ for p. We may say that the group characters ζ_μ of $\mathfrak{G}$ are built up by the modular characters φ_ν, and it is possible to obtain a deeper insight into the nature of the ordinary group characters by the use of the modular characters and their properties. However, it is disturbing that we have to restrict ourselves to p-regular elements. In this paper, we plan to overcome this difficulty. The value $\zeta_\mu(G)$ for elements G of an order divisible by p will be expressed by the modular characters of certain subgroups $\mathfrak{N}_i$ of $\mathfrak{G}$. The corresponding generalized decomposition numbers $d^i_{\mu\nu}$ will not necessarily be rational, but they are integers of a cyclotomic field of an order p^α. The definition of these numbers $d^i_{\mu\nu}$, and the formulae generalizing (*) are given in §1. The numbers $d^i_{\mu\nu}$ can be arranged in the form of a square matrix $\mathbf{D}$ which is non-degenerate, and, apart from the arrangements of the rows and columns, the $d^i_{\mu\nu}$ are invariants of the group $\mathfrak{G}$. The matrix of the group characters of $\mathfrak{G}$ can be written as the product of $\mathbf{D}$ and a matrix A which breaks up completely[2] into the matrices of the modular group characters of $\mathfrak{N}_0$, $\mathfrak{N}_1$, $\mathfrak{N}_2$, $\cdots$, if the rows and columns in all these matrices are suitably arranged. If the modular characters of the groups $\mathfrak{N}_i$ are known, the product of any two columns of $\mathbf{D}$

* Presented to the American Mathematical Society on April 11, 1941.

[1] R. Brauer and C. Nesbitt, On the modular representations of groups of finite order, University of Toronto Studies, Math. Series No. 4, 1937. R. Brauer and C. Nesbitt, On the modular characters of groups. I refer to these two papers by BN 1 and BN 2. The introduction of BN 2 contains a short summary of most of the methods and results of BN 1.

[2] This means that the matrix A contains in its main diagonal the matrices of the modular group characters of $\mathfrak{N}_0$, $\mathfrak{N}_1$, $\cdots$, and zero matrices outside the main diagonal.

can be formed; the Cartan invariants of the groups $\mathfrak{N}_i$ appear as the values of these products (§3). Further, some congruences (mod p) for the decomposition numbers are given which will be of fundamental importance for a later paper.

With every column of $\mathbf{D}$, all algebraically conjugate columns appear in $\mathbf{D}$. In this manner, the columns of $\mathbf{D}$ are distributed in "families" of conjugate columns. The number N of such families and the numbers $w_1, w_2, \cdots, w_N$ of members of the individual families are determined in §§6, 7. The numbers obtained are the same as if we distribute the characters ζ_μ into "families" of p-conjugate characters where two characters are said to be *p-conjugate*, if they can be transformed into each other by a change of a primitive $p^{a\text{th}}$ root of unity. Finally, we also obtain these numbers N, w_i, when we distribute the classes of conjugate elements into "families" in a certain manner which can be described in terms of abstract group theory.

Notation: $\mathfrak{G}$ denotes a group of finite order g, p is a fixed rational prime number, and we set $g = p^a g'$ with $(p, g') = 1$. A *p-regular* element of $\mathfrak{G}$ is an element whose order is prime to p; the other elements are said to be *p-singular*. Similarly, we denote the classes of conjugate elements as p-regular or p-singular according as the elements of the class are p-regular or p-singular. The normalizor $\mathfrak{N}(G)$ of an element G consists of those elements of $\mathfrak{G}$ which commute with G; the order of $\mathfrak{N}(G)$ will be denoted by $n(G)$. When we speak of a representation or a character of $\mathfrak{G}$ without further attribute, we mean a representation in the field of all complex numbers and the corresponding character. In the case of a representation in a field of characteristic p, we always add the word "modular".

1. Definition of the generalized decomposition numbers

We consider a group $\mathfrak{G}$ of finite order $g = p^a g'$ where p is a prime number and $(g', p) = 1$. Let $P_0 = 1, P_1, P_2, \cdots, P_h$ be a system of elements whose orders are powers of p, such that they all lie in different classes of conjugate elements, but that every element of an order p^α $(\alpha = 0, 1, 2, \cdots)$ is conjugate to one of them. Of course, the P_i can be taken from a fixed Sylow subgroup of order p^a. Every p-singular class of $\mathfrak{G}$ contains an element of the form P_iV where i is uniquely determined by the class and where V is a p-regular element of the normalizor $\mathfrak{N}(P_i)$ of P_i. If W is a p-regular element of $\mathfrak{N}(P_j)$ and P_iV and P_jW are conjugate, say

$$G^{-1}P_iVG = P_jW,$$

then by raising this equation to suitable powers, we find

$$G^{-1}P_iG = P_j, \qquad G^{-1}VG = W.$$

Hence $i = j$, G lies in $\mathfrak{N}(P_i)$, and V and W are conjugate in $\mathfrak{N}(P_i)$. Conversely, these conditions imply that P_iV and P_jW are conjugate in $\mathfrak{G}$. In order to obtain a complete system of representatives for the p-singular classes of $\mathfrak{G}$, we have to form P_iV where $i = 0, 1, 2, \cdots, h$, and, for each i, the element V ranges

over a complete system of representatives of the p-regular classes of the group $\mathfrak{N}(P_i)$. Let k_i denote the number of these classes.

Consider a representation $\mathfrak{F}$ of $\mathfrak{G}$. The matrix $\mathfrak{F}(P_i)$ representing a fixed element P_i can be assumed to be of the form

$$\mathfrak{F}(P_i) = \begin{pmatrix} \epsilon_1 I_1 & 0 & \cdots & 0 \\ 0 & \epsilon_2 I_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \epsilon_l I_l \end{pmatrix},$$

where $I_1, I_2, \cdots, I_l$ are unit matrices, and $\epsilon_1, \epsilon_2, \cdots, \epsilon_l$ are distinct $p^{\alpha\text{th}}$ roots of unity; for p^α we may take the order of P_i. The matrix representing an element V_i of $\mathfrak{N}(P_i)$ then breaks up in a corresponding manner,

$$\mathfrak{F}(V_i) = \begin{pmatrix} N^{(1)} & 0 & \cdots & 0 \\ 0 & N^{(2)} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & N^{(l)} \end{pmatrix}.$$

The trace tr $(\mathfrak{F}(P_iV_i))$ is given by

$$\operatorname{tr}(\mathfrak{F}(P_iV_i)) = \epsilon_1 \operatorname{tr}(N^{(1)}) + \epsilon_2 \operatorname{tr}(N^{(2)}) + \cdots + \epsilon_l \operatorname{tr}(N^{(l)}).$$

If V_i ranges over all elements of $\mathfrak{N}(P_i)$, then $N^{(\lambda)}$, for a fixed λ, will form a representation of $\mathfrak{N}(P_i)$. The trace tr $(N^{(\lambda)})$ can therefore be expressed as a linear combination of the irreducible characters of $\mathfrak{N}(P_i)$. If V_i is p-regular, these characters again can be expressed by the irreducible modular characters of $\mathfrak{N}(P_i)$.

Let $\varphi_1^i, \varphi_2^i, \cdots$ be the distinct absolutely irreducible modular characters of $\mathfrak{N}(P_i)$ for p; we have k_i of them.[3] Thus we obtain a formula

$$\operatorname{tr}(\mathfrak{F}(P_iV_i)) = \sum d_\nu^i \varphi_\nu^i(V_i)$$

where d_ν^i is an integer of the field of the $p^{\alpha\text{th}}$ roots of unity; it is independent of V_i.

If $\mathfrak{F}_1, \mathfrak{F}_2, \cdots$ are the distinct irreducible representations of $\mathfrak{G}$, and $\zeta_1, \zeta_2, \cdots$ the corresponding characters, we therefore have formulae

$$(1) \qquad \zeta_\mu(P_iV_i) = \sum_{\nu=1}^{k_i} d_{\mu\nu}^i \varphi_\nu^i(V_i) \qquad (V_i \text{ in } \mathfrak{N}(P_i), p\text{-regular}).$$

We denote the $d_{\mu\nu}^i$ as the *generalized decomposition numbers* of $\mathfrak{G}$. For $i = 0$, we have $P_0 = 1$, $\mathfrak{N}(P_0) = \mathfrak{G}$, and the $d_{\mu\nu}^0$ are identical with the ordinary decomposition numbers $d_{\mu\nu}$ of $\mathfrak{G}$.[4] In any case, $d_{\mu\nu}^i$ is an integer of the field of the $p^{\alpha_i\text{th}}$ roots of unity where p^{α_i} is the order of P_i.

Let Φ_ν^i be the indecomposable modular character which corresponds to φ_ν^i,

---

[3] BN 1, Theorem III.—BN 2, §8.

[4] BN 1, p. 7.—BN 2, §4.

($\nu = 1, 2, \cdots, k_i$). If $n(P_i)$ is the order of $\mathfrak{N}(P_i)$, then it follows from the orthogonality relations for modular group characters[5] that

$$d^i_{\mu\nu} = \frac{1}{n(P_i)} \sum_{V \text{ in } \mathfrak{N}(P_i)}' \zeta_\mu(P_i V)\Phi^i_\nu(V^{-1}), \tag{2}$$

where the dash indicates that V ranges over the p-regular elements of $\mathfrak{N}(P_i)$ only. We arrange these numbers $d^i_{\mu\nu}$ for a fixed i in form of a matrix $D^i = (d^i_{\mu\nu})$ with μ as row index and ν as column index, and set

$$\mathbf{D} = (D^0, D^1, \cdots, D^h). \tag{3}$$

Any column of $\mathbf{D}$ will be denoted as a *d-column of* $\mathfrak{G}$; each of them is given by a pair (i, ν). It follows that the number of such columns is equal to the sum of all k_i, i.e. equal to the full number k of classes of conjugate elements of $\mathfrak{G}$. Hence $\mathbf{D}$ is a square matrix of the same degree k as the matrix Z of the group characters ζ_μ of $\mathfrak{G}$. According to (1), we have a formula $Z = \mathbf{D}A$ where A is a square matrix. Since Z is non-singular, so is $\mathbf{D}$.

THEOREM 1: *The matrix* $\mathbf{D}$ *of the generalized decomposition numbers of* $\mathfrak{G}$ *is non-singular. The matrix of the group characters of* $\mathfrak{G}$ *has the form* $\mathbf{D}A$ *where the matrix* A *breaks up completely*[6] *into the matrices of the modular group characters of* $\mathfrak{N}(P_0), \mathfrak{N}(P_1), \cdots, \mathfrak{N}(P_h)$, *provided that the rows and columns of the matrices are suitably arranged.*

If $\mathbf{D}$ and the modular group characters of all $\mathfrak{N}(P_i)$ are known, then the ordinary group characters ζ_μ of $\mathfrak{G}$ can be found from (1).

2. Change of P_i

For the definition of the generalized decomposition numbers $d^i_{\mu\nu}$, a special system of elements P_i of $\mathfrak{G}$ has been used. We now have to see how the $d^i_{\mu\nu}$ are affected by an admissible change in the choice of P_i, that is when we replace P_i by a conjugate element $G^{-1}P_iG = P_i^*$. Of course, $G^{-1}\mathfrak{N}(P_i)G = \mathfrak{N}(P_i^*)$. If $V \to \mathfrak{F}(V)$, (V in $\mathfrak{N}(P_i)$), is a representation of $\mathfrak{N}(P_i)$, then $V^* \to \mathfrak{F}(GV^*G^{-1})$, V^* in $\mathfrak{N}(P_i^*)$, is a representation of $\mathfrak{N}(P_i^*)$. If $\chi(V)$ is the character of $\mathfrak{F}$, we denote the character of the new representation by χ^G, i.e.

$$\chi^G(V^*) = \chi(GV^*G^{-1}), \qquad V^* \text{ in } \mathfrak{N}(P_i^*). \tag{4}$$

Then

$$(\varphi^i_1)^G, (\varphi^i_2)^G, \cdots, (\varphi^i_{k_i})^G$$

is a complete system of irreducible modular characters of $\mathfrak{N}(P_i^*)$; the corresponding indecomposable characters are

$$(\Phi^i_1)^G, (\Phi^i_2)^G, \cdots, (\Phi^i_{k_i})^G.$$

Since $\zeta_\mu(P_iV) = \zeta_\mu(G^{-1}P_iVG)$, we obtain easily from (2)

[5] BN 1, Theorem IV.—BN 2, (20).

[6] Cf. footnote 2.

THEOREM 2: *If P_i is replaced by $G^{-1}P_iG = P_i^*$, then $d^i_{\mu\nu} = d^{i*}_{\mu\tau}$, where $d^{i*}_{\mu\tau}$ are the decomposition numbers corresponding to this new choice, and where $(\varphi^i_\nu)^G$ is taken as the τ^{th} character of $\mathfrak{N}(P_i^*)$.*

We see that the only possible change is a permutation of the columns of D^i.

THEOREM 3: *The generalized decomposition numbers are invariants of $\mathfrak{G}$.*

3. The orthogonality relations

We form the "unitary product" of two columns of $\mathbf{D}$. According to (2), we have

$$(5)\quad \sum_{\mu=1}^{k} d^i_{\mu\nu}\bar{d}^j_{\mu\rho} = \frac{1}{n(P_i)n(P_j)} \sum_{V \text{ in } \mathfrak{N}(P_i)}' \sum_{W \text{ in } \mathfrak{N}(P_j)}' \Phi^i_\nu(V^{-1})\Phi^j_\rho(W) \sum_\mu \zeta_\mu(P_iV)\overline{\zeta_\mu(P_jW)}.$$

The sum over μ on the right hand side vanishes, if P_iV and P_jW are not conjugate in $\mathfrak{G}$; in the other case its value is $n(P_iV)$. It follows that, for $i \neq j$, the whole expression vanishes. For $i = j$, only the $n(P_i)/n(P_iV)$ elements W are to be taken into account which, in $\mathfrak{N}(P_i)$, are conjugate to V. Hence

$$\sum_{\mu=1}^{k} d^i_{\mu\nu}\bar{d}^i_{\mu\rho} = \frac{1}{n(P_i)} \sum_{V \text{ in } \mathfrak{N}(P_i)}' \Phi^i_\nu(V^{-1})\Phi^i_\rho(V).$$

The right hand side can easily be evaluated.[7] We thus obtain

THEOREM 4: *The generalized decomposition numbers satisfy the equations*

$$(6)\qquad \sum_{\mu=1}^{k} d^i_{\mu\nu}\bar{d}^j_{\mu\rho} = \begin{cases} 0 & \text{for } i \neq j \\ c^i_{\nu\rho} & \text{for } i = j \end{cases}$$ [8]

where $c^i_{\nu\rho}$ is the Cartan invariant of $\mathfrak{N}(P_i)$ corresponding to the modular characters $\varphi^i_\nu, \varphi^i_\rho$.

On multiplying $\mathbf{D}'$ with the conjugate complex matrix $\bar{\mathbf{D}}$ and forming the determinant of the product, we obtain the product of all the determinants $|c^i_{\kappa\lambda}|$, $i = 0, 1, 2, \cdots, h$. All these $|c^i_{\kappa\lambda}|$ are powers of p.[9] With every d-column, the conjugate complex column also appears as a d-column, as is easily seen. Hence $|\bar{\mathbf{D}}| = \pm|\mathbf{D}|$. Since the determinant of the matrix A in theorem 1 is prime to p,[10] we have

THEOREM 5: *The square of the determinant of $\mathbf{D}$ is of the form $\pm p^m$, m a rational integer > 0. The determinant of the second factor A in theorem 1 is prime to p.*

It is not difficult to determine the exact value of the determinant of $\mathbf{D}$.

---

[7] BN 1, p. 15.—BN 2, (22).

[8] It should be noted that the $d^i_{\mu\nu}$ are not the ordinary decomposition numbers of $\mathfrak{N}(P_i)$ though they satisfy exactly the same relations.

[9] Cf. R. Brauer, On the Cartan invariants of groups of finite order, Annals of Math. 42, p. 53, 1941.

[10] BN 2, §8.—In BN 1, it is shown that the determinant $|X|$ of the matrix X of the modular group characters of a group is prime to a fixed prime ideal divisor $\mathfrak{p}$ of p, cf. BN 1, (26). The proof given in BN 1, p. 14 for the fact that $|X|$ is prime to p, is not correct. However, this result follows from BN 1, (26), since $|X|^2$ is a rational integer, cf. BN 1, (29).

4. p-conjugate characters

If we replace a primitive g^{th} root of unity ϵ_g by another one ϵ_g^λ, $(\lambda, g) = 1$, then every character ζ_μ is transformed into a conjugate character ζ_σ. Choose now $\lambda \equiv 1 \pmod{g'}$ so that the substitution $\epsilon_g \to \epsilon_g^\lambda$ amounts to an interchange of the $p^{a\text{th}}$ roots of unity, such that the g'^{th} roots of unity remain unaltered. In this case, we say that the two conjugate characters ζ_μ and ζ_σ are *p-conjugate.* All the characters are distributed into *families of p-conjugate characters.*

If ζ_μ and ζ_σ are p-conjugate, then $\zeta_\mu(G) = \zeta_\sigma(G)$ for p-regular elements G. It follows that ζ_μ and ζ_σ have the same modular constituents (for p). Hence

THEOREM 6: *Two p-conjugate characters ζ_μ and ζ_σ have the same modular constituents; they lie in the same block of characters (for p).*

5. The decomposition numbers corresponding to a block of characters

Let B be a block[11] of characters of $\mathfrak{G}$ (for p). We consider a sum analogous to (6) but where μ ranges only over the values for which ζ_μ belongs to B. We shall say for short that these are the *indices in* B. Similar to (5) we have

$$\sum_{\mu \text{ in B}} d^i_{\mu\nu} \bar{d}^j_{\mu\rho} = \sum_V{}' \sum_W{}' \frac{1}{n(P_i)} \frac{1}{n(P_j)} \Phi^i_\nu(V^{-1}) \Phi^j_\rho(W) \sum_{\mu \text{ in B}} \zeta_\mu(P_i V) \overline{\zeta_\mu(P_j W)}.$$

As the $d^i_{\mu\nu}$, the whole sum is an integer of the field of the $p^{a\text{th}}$ roots of unity. But with every ζ_μ all its p-conjugates appear, and the expression on the right hand side shows that the sum is a rational integer.

Now collect the terms for which V lies in a fixed class of $\mathfrak{N}(P_i)$ and W in a fixed class of $\mathfrak{N}(P_j)$. Since these classes consist of $n(P_i)/n(P_i V)$ and of $n(P_j)/n(P_j W)$ elements respectively, and all the corresponding terms have the same value, we find

$$\begin{aligned} &\sum_\nu d^i_{\mu\nu} \bar{d}^j_{\mu\rho} \\ (7) \qquad &= \sum_V{}'' \sum_W{}'' [\Phi^i_\nu(V^{-1})/n(P_i V)][\Phi^j_\rho(W)/n(P_j W)] \sum_{\mu \text{ in B}} \zeta_\mu(P_i V)\overline{\zeta_\mu(P_j W)} \end{aligned}$$

where V and W range over certain elements of $\mathfrak{N}(P_i)$ and $\mathfrak{N}(P_j)$ respectively. The numbers in the square brackets are p-integers.[12] If $\mathfrak{p}$ is a prime ideal divisor of p in the field generated by the characters ζ_μ, then

$$\zeta_\mu(P_i V) \equiv \zeta_\mu(V) \pmod{\mathfrak{p}}.$$

On the other hand, if $j > 0$ and, therefore, $P_j W$ p-singular, we have

$$\sum_{\mu \text{ in B}} \zeta_\mu(V)\overline{\zeta_\mu(P_j W)} = 0.^{13}$$

[11] BN 1, §§6, 7.—BN 2, §9.

[12] BN 1, theorem V.—This theorem is a consequence of BN 2, (16) and (17).

[13] BN 1, theorem VIII.—This can also be seen from BN 2, (28), and the formulae (1) and (6) of the present paper.

Hence for $j > 0$ the sum (7) is divisible by $\mathfrak{p}$, and since it is rational, it is divisible by p.

THEOREM 7: *If* $P_j \neq 1$, *i.e. if* $j > 0$, *then for every block* B

$$(8) \qquad \sum_{\mu \text{ in } B} d^i_{\mu\nu} \bar{d}^i_{\mu\rho} \equiv 0 \pmod{\mathfrak{p}};$$

the left side is a rational integer.

Assume that the blocks $B = B_\lambda$ consists of the ordinary characters $\zeta_1, \zeta_2, \cdots, \zeta_x$ and the modular characters $\varphi_1, \varphi_2, \cdots, \varphi_y$, $(\varphi_\nu = \varphi^0_\nu)$ and form the matrix $D_\lambda = (d_{\rho\sigma}) = (d^0_{\rho\sigma})$, $(\rho = 1, 2, \cdots, x; \sigma = 1, 2, \cdots, y)$. The matrix D breaks up completely into $D_1, D_2, \cdots$ corresponding to the different blocks.[14] From (6), it follows that

$$(8^*) \qquad \sum_{\mu \text{ in } B} d^i_{\mu\nu} d^0_{\mu\rho} = \begin{cases} c^0_{\nu\rho} = c_{\nu\rho}, & \text{if } i = 0, \text{ and } \varphi_\nu, \varphi_\rho \text{ both belong to B.} \\ 0, & \text{in all other cases.} \end{cases}$$

This supplements (8).

If $\zeta_1, \zeta_2, \cdots, \zeta_x$ belong to w different families of p-conjugate characters, then we arrange the ζ_μ so that $\zeta_1, \zeta_2, \cdots, \zeta_w$ all belong to different families. Assume that the family of ζ_ν consists of r_ν characters and set

$$(9) \qquad \tilde{D}_\lambda = (d_{\rho\sigma}) \qquad (\rho = 1, 2, \cdots, w; \sigma = 1, 2, \cdots, y).$$

If ζ_μ and ζ_ρ are p-conjugate, then $d_{\mu\sigma} = d_{\rho\sigma}$ by theorem 6. Therefore, D_λ has the same rows as $\tilde{D}_\lambda$, but the μ^{th} row of $\tilde{D}_\lambda$ appears r_μ times in D_λ. From (8*) and (1), it follows that

$$\sum_{\mu=1}^{x} d_{\mu\nu} \zeta_\mu(S) = 0 \qquad (\text{for all } p\text{-singular } S \text{ of } \mathfrak{G}).$$

Here, p-conjugate characters have the same coefficients. If we denote by $\tilde{\zeta}_\mu$ the sum of all characters which are p-conjugate to ζ_μ (including ζ_μ), the equation can be written in the form

$$(10) \qquad \sum_{\mu=1}^{w} d_{\mu\nu} \tilde{\zeta}_\mu(S) = 0 \qquad (\text{for all } p\text{-singular } S \text{ of } \mathfrak{G}).$$

We set $d_\mu = \sum_\nu d_{\mu\nu}\omega_\nu$ for any fixed numbers $\omega_1, \cdots, \omega_y$. Then

$$(10a) \qquad \sum_{\mu=1}^{w} d_\mu \tilde{\zeta}_\mu(S) = 0 \qquad (\text{for all } p\text{-singular } S \text{ of } \mathfrak{G}).$$

Every character $\zeta_\mu(G)$ of $\mathfrak{G}$ may be considered as a character of the p-Sylow-subgroup $\mathfrak{P}$ of $\mathfrak{G}$, when we take G as an element P of $\mathfrak{P}$. If, in this sense, ζ_μ contains the 1-character [1] of $\mathfrak{P}$ q_μ times, the same will hold for the p-conjugate characters and, therefore, $\tilde{\zeta}_\mu$ will contain [1] exactly $r_\mu q_\mu$ times. The expression $\chi(P) = \sum d_\mu \tilde{\zeta}_\mu(P)$ is a linear combination of the characters of $\mathfrak{P}$, and [1] appears

[14] BN 1, theorem VII.—BN 2, (28).

in it with the coefficient $\sum d_\mu r_\mu q_\mu$. But, from (10a), $\chi(P) = 0$ for $P \neq 1$. If z_μ is the degree of ζ_μ, then $\chi(1) = \sum r_\mu d_\mu z_\mu$. The orthogonality relations for ordinary group characters give $(1/p^a) \sum r_\mu d_\mu z_\mu$ as the coefficient of [1] in χ, and hence

$$(11) \qquad p^a \sum_{\mu=1}^{w} r_\mu d_\mu q_\mu = \sum_{\mu=1}^{w} r_\mu d_\mu z_\mu .$$

If only one of the numbers d_μ is different from 0, (11) becomes $p^a q_\mu = z_\mu$, and then the character ζ_μ is of the highest kind.[15] If we exclude this case, it follows that it is impossible to determine $\omega_1, \omega_2, \cdots, \omega_y$ so that only one of the d_μ does not vanish. This implies that the rank of $\tilde{D}_\lambda$ is smaller than w. But $\tilde{D}_\lambda$ has the same rank as D_λ, and this rank is y.[16] Hence

THEOREM 8: *Every block* B *which is not of the highest kind contains more families of ordinary characters than it contains modular characters*: $w > y$.[17]

Each relation (10) must contain at least two $\tilde{\zeta}_\mu$. This gives

THEOREM 9: *If the block* B *is not of the highest kind, then each of its modular constituents appears in at least two characters* ζ_μ *of* B *which are not p-conjugate.*

From (10) it also follows that

$$(12) \qquad \sum_{\mu=1}^{w} \zeta_\mu(R)\tilde{\zeta}_\mu(S) = 0$$

for any p-regular R and any p-regular S.

The blocks of highest kind consist of exactly one ζ_μ whose degree is divisible by p^a, and each such ζ_μ forms a block of highest kind.[18] Since such a ζ_μ vanishes for all p-singular elements, (2) gives

THEOREM 10: *If* ζ_μ *forms a block of highest kind, then* $d^i_{\mu\nu} = 0$ *for all* $i > 0$ *and all* ν.

6. The permutation lemma

We now derive a simple lemma which we shall need. Consider a matrix $M = (m_{ij})$ with u rows and v columns. Every permutation A of the rows of M can be effected by left-multiplication of M with a suitable "permutation matrix" P_A of degree u which in every row and in every column has one coefficient 1 and $m - 1$ coefficients 0. Similarly, every permutation B of the columns of M can be effected by right-multiplication of M with a suitable permutation matrix Q_B of degree v. We prove

LEMMA 1: *Let M be a non-singular matrix. If there exists a permutation A of the rows of M and a permutation B of the columns of M such that both carry M into the same matrix, then the cycles of the permutation A have the same lengths as those of B. In particular, A and B have the same number of cycles.*

[15] BN 2, theorem 1.
[16] BN 1, p. 21.—BN 2, (29), (15) and (14).
[17] This improves the inequality $x > y$ given in BN 2.
[18] Cf. footnote 15.

PROOF: According to the assumption, we have

$$P_A M = M Q_B. \tag{13}$$

Since M is non-singular, P_A and Q_B have the same characteristic roots. To each cycle of length r of A, there correspond the r r^{th} roots of unity as characteristic roots of P_A.[19] On comparing the roots of P_A and Q_B, starting with the maximal r, we readily obtain lemma 1. Similarly, we prove

LEMMA 2: *Let M be a rectangular matrix whose columns are linearly independent, and assume that there exists a permutation A of the rows and a permutation B of the columns of M which both carry M into the same matrix. If B has a cycle of length r then A has a cycle whose length is divisible by r.*

PROOF: Again, (13) holds. Schur's lemma shows that Q_B is a constituent of P_A so that the characteristic roots of Q_B appear among those of P_A. On comparing these roots, we obtain lemma 2. We also have the result

LEMMA 3: *Let M be a non-singular matrix of degree m, and let $\mathfrak{A}$ and $\mathfrak{B}$ be two permutation groups of degree m which are both homomorphic to the same group $\mathfrak{T}$. If for every T in $\mathfrak{T}$, the corresponding element A_T of $\mathfrak{A}$, applied to the rows of M, and the corresponding element B_T of $\mathfrak{B}$, applied to the columns of M, both carry M into the same matrix, then the number of systems of transitivity is the same for $\mathfrak{A}$ and for $\mathfrak{B}$.*

PROOF: Again (13) will hold for corresponding $A = A_T$ and $B = B_T$. All we have to show is that the number $\tau_{\mathfrak{A}}$ of systems of transitivity is an invariant, if $\mathfrak{A}$ is interpreted as a group of linear transformations, and similarity transformations of $\mathfrak{A}$ are performed. But this follows from the fact that $\tau_{\mathfrak{A}}$ is the number of 1-constituents of the linear group $\mathfrak{A}$.

7. Families of characters, classes, and d-columns

The results of §6 can be applied when M is the matrix of group characters of $\mathfrak{G}$. We may construct corresponding permutations A and B in the following manner. Let ϵ_g be a primitive g^{th} root of unity, and let λ be a rational integer which is prime to g. The substitution $T_\lambda : \epsilon_g \to \epsilon_g^\lambda$ carries every character χ of $\mathfrak{G}$ into a conjugate character $\chi^{(\lambda)}$, we have

$$\chi^{(\lambda)}(G) = \chi(G^\lambda). \tag{14}$$

On the other hand, the substitution $G \to G^\lambda$ carries every class of conjugate elements $\mathfrak{C}$ into a new class $\mathfrak{C}^{(\lambda)}$. Then (14) shows that the value of χ for $\mathfrak{C}^{(\lambda)}$ is the same as the value of $\chi^{(\lambda)}$ for $\mathfrak{C}$. Hence the permutation $A: \chi \to \chi^{(\lambda)}$ of the rows of M, and the permutation $B: \mathfrak{C} \to \mathfrak{C}^{(\lambda)}$ of the columns both carry M into the same matrix.

We are interested in the case that $\lambda \equiv 1 \pmod{g'}$. Then χ and $\chi^{(\lambda)}$ belong to

19 A modification is necessary, if the underlying field is modular, but the lemma remains valid. The same is true for lemma 2 and lemma 3. We shall use the lemmas only in the case of a non-modular field.

the same family of p-conjugate classes (§4). We shall also say that the classes $\mathfrak{C}$ and $\mathfrak{C}^{(\lambda)}$ belong to the same family of classes ($\lambda \equiv 1 \pmod{g'}$). If $\mathfrak{C}$ contains the element $P_i V$ (V in $\mathfrak{N}(P_i)$, p-regular), then $\mathfrak{C}^{(\lambda)}$ contains $P_i^\lambda V$ in this case.

Before formulating the results, we also consider the matrix $\mathbf{D}$, (3). With every column $d^i_{\mu\nu}$, all the algebraically conjugate columns will appear. Indeed, the substitution T_λ, ($\lambda \equiv 1 \pmod{g'}$) results in the replacing of P_i in (2) by P_i^λ, and on account of Theorem 2, this new column can again be expressed in the form $d^j_{\mu\tau}$. The d-columns thus appear distributed into families of algebraically conjugate d-columns. The effect of the substitution T_λ on $\mathbf{D}$ then consists of a permutation B^* of the columns; the members of each family are interchanged among themselves.

On the other hand, the effect of T_λ on $\mathbf{D}$ can also be described by the permutation $A: \chi \rightarrow \chi^{(\lambda)}$ of the rows of $\mathbf{D}$ as follows from (2). Hence the assumptions of lemma 1 are also satisfied for $M = \mathbf{D}$ and the permutations A and B^*.

Let $\mathfrak{T}$ be the group of all substitutions T_λ with $\lambda \equiv 1 \pmod{g'}$. We then have homomorphic groups $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{B}^*$ consisting of the A, the B, and the B^* respectively. In each of the three cases, a system of transitivity corresponds exactly to a family (of characters, classes, or d-columns). Hence lemma 3 gives

THEOREM 11: *The number of distinct families is the same for each of the three kinds of families: Families of p-conjugate characters, families of classes, and families of conjugate d-columns.*[20]

Next let p be an odd prime. Thus $\mathfrak{T}$ is cyclic, and a primitive element is obtained by taking for λ a primitive root $\pmod{p^a}$, ($\lambda \equiv 1 \pmod{g'}$). For this T_λ, the lengths of the cycles of the permutation A are the numbers of members belonging to the different families of characters. A similar statement holds for B and B^*. Lemma 1 now yields

THEOREM 12: *Let* $p \neq 2$. *If the different families of characters contain* $r_1, r_2, \cdots, r_f$ *members respectively, if the different families of classes contain* $s_1, s_2, \cdots, s_f$ *respectively, and if the different families of d-columns contain* $t_1, t_2, \cdots, t_f$ *members respectively, then the three sets* $(r_1, r_2, \cdots, r_f)$, $(s_1, s_2, \cdots, s_f)$, $(t_1, t_2, \cdots, t_f)$ *are identical apart from the arrangement.*[21]

Remark: The k_0 p-regular classes form each a family of its own, $s_\kappa = 1$. Similarly, the k_0 d-columns of $D^0 = D$ form each a family of its own, $t_\kappa = 1$.

There is another case in which lemma 1 and lemma 3 can be applied. Every automorphism of $\mathfrak{G}$ permutes the characters, the classes, and the d-columns, and again the assumptions of the lemmas are satisfied. It seems unnecessary to formulate the results explicitly.

UNIVERSITY OF TORONTO,
TORONTO, CANADA.

[20] For the first two kinds of families, compare the similar statement and proof in W. Burnside, Theory of groups of finite order, 2nd. ed., Cambridge 1911, p. 315, theorem VI.

[21] For $p = 2$ this will hold, if G does not contain elements of order 8.

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

INVESTIGATIONS ON GROUP CHARACTERS*

BY RICHARD BRAUER

(Received November 19, 1940)

Introduction

Let $\mathfrak{G}$ be a finite group of order g. It is well known that the distinct irreducible representations $\mathfrak{Z}_1, \mathfrak{Z}_2, \cdots, \mathfrak{Z}_k$ of $\mathfrak{G}$ in the field of complex numbers can be so chosen that their coefficients belong to an algebraic number field Ω of finite degree. Furthermore, if p is a fixed rational prime number, we may assume that the coefficients are p-integers, i.e. are of the form α/β where α and β are integers of Ω and β is prime to p. Let $\mathfrak{P}$ be a fixed prime ideal divisor of p in Ω. If every coefficient of $\mathfrak{Z}_\kappa$ is replaced by its residue class (mod $\mathfrak{P}$), then we obtain a modular representation $\bar{\mathfrak{Z}}_\kappa$ with coefficients in a field of characteristic p.

It was shown by Dickson and Speiser[1] that the ordinary theory of group characters remains valid for modular group characters, if the prime p does not divide the group order g. Every modular representation then is completely reducible; all the distinct, absolutely irreducible modular representations are given by $\bar{\mathfrak{Z}}_1, \bar{\mathfrak{Z}}_2, \cdots, \bar{\mathfrak{Z}}_k$.

In a recent paper,[2] C. Nesbitt and the author obtained results which may be considered as generalizations of these theorems. Let p be any rational prime, and assume that p^a is the highest power of p which divides g, say

$$g = p^a g', \qquad (p, g') = 1.$$

We then considered representations $\mathfrak{Z}_\kappa$ whose degree is divisible by p^a. It was shown that $\bar{\mathfrak{Z}}_\kappa$ is absolutely irreducible as a modular representation. Whenever $\bar{\mathfrak{Z}}_\kappa$ appears as a constituent of a modular representation $\mathfrak{F}$, then $\mathfrak{F}$ breaks up completely into $\bar{\mathfrak{Z}}_\kappa$ and another constituent $\mathfrak{A}$ (reducible or irreducible)

$$\mathfrak{F} \sim \begin{pmatrix} \bar{\mathfrak{Z}}_\kappa & 0 \\ 0 & \mathfrak{A} \end{pmatrix}.$$

None of the representations $\bar{\mathfrak{Z}}_\lambda$ for $\lambda \neq \kappa$ contains $\bar{\mathfrak{Z}}_\kappa$ as a constituent. Since every irreducible modular representation appears as a constituent of at least one of the representations $\bar{\mathfrak{Z}}_1, \bar{\mathfrak{Z}}_2, \cdots, \bar{\mathfrak{Z}}_k$, this, (in the case $a = 0$, i.e. $g \not\equiv 0$ (mod p)), actually yields the theorem of Dickson and Speiser.

* Presented to the American Mathematical Society on September 5, 1941.

[1] L. E. Dickson, Trans. Am. Math. Soc. **3**, p. 285, 1902. A. Speiser, *Theorie der Gruppen von endlicher Ordnung* 3rd ed. Berlin 1937, §71.

[2] *On the modular characters of groups*, Ann. of Math., **42**, p. 556, 1941. I refer to this paper as BN.

In this paper, I study representations $\mathfrak{Z}_\kappa$ whose degree z_κ is divisible by p^{a-1} but not by p^a. If the order g of $\mathfrak{G}$ is divisible by p to the first power only, then every representation $\mathfrak{Z}_\kappa$ is either of this type, or of the "highest" type, $z_\kappa \equiv 0 \pmod{p^a}$, which was studied in the former paper. Our results will enable us to derive a great number of properties of the characters of groups $\mathfrak{G}$ of such an order $g = pg'$; these properties form a powerful weapon in the investigation of these groups.[3]

Our first result concerning representations $\mathfrak{Z}_\kappa$ of a degree $z_\kappa \equiv 0 \pmod{p^{a-1}}$ is that the degree of all those representations $\mathfrak{Z}_\lambda$ which belong to the same block[4] B as $\mathfrak{Z}_\kappa$, is also divisible by p^{a-1}. I mention here the following application of this theorem: The only simple group of an order $4p^aq^b$, (p, q primes) with $a \leqq 2$ is the simple group of order 60; the only simple groups of an order $3p^aq^b$ (p, q primes) with $a \leqq 2$ are the simple groups of order 60 and 168. (§9.)

With the block B containing a representation $\mathfrak{Z}_k$ of degree $z_\kappa \equiv 0 \pmod{p^{a-1}}$, $z_\kappa \not\equiv 0 \pmod{p^a}$, we associate a linear graph. Every vertex V_λ corresponds to a family of p-conjugate characters[5] ζ_λ, ζ'_λ, $\cdots$ of B, every edge S_μ to a modular character φ_μ of B, and the edge S_μ contains V_λ, if φ_μ is a modular constituent of ζ_λ (and so of all its p-conjugates). It will be shown that every S_λ contains only two vertices, further that the complex actually is a tree T. Since we also have the result that φ_μ never appears with higher multiplicity than 1 in an ordinary irreducible character, the tree describes the complete structure of the block B, if at every vertex V_λ the number r_λ of characters in the family of ζ_λ is indicated. If T and these numbers r_λ are given, the decomposition numbers and hence the Cartan invariants corresponding to the block can be easily obtained. Of course, there exist two vertices of T which lie on only one edge. Consequently, the block B contains at least two characters which are not p-conjugate, and which remain irreducible, when considered as modular characters.

If $\mathfrak{Z}_\lambda$ is any irreducible representation of $\mathfrak{G}$, if z_λ is its degree and r_λ the number of its p-conjugates, then we can show that $r_\lambda z_\lambda \not\equiv 0 \pmod{p^{a+1}}$.[6] If $r_\lambda z_\lambda \equiv 0 \pmod{p^a}$, then $z_\lambda \equiv 0 \pmod{p^a}$, i.e. $\mathfrak{Z}_\lambda$ is of the highest kind.

In the case $z_\lambda \equiv 0 \pmod{p^{a-1}}$, $z_\lambda \not\equiv 0 \pmod{p^a}$, it follows that r_λ divides $p - 1$. Besides, we have the relation

$$r_1 r_2 \cdots r_w \left(\frac{1}{r_1} + \frac{1}{r_2} + \cdots + \frac{1}{r_w}\right) = p$$

for the numbers r_λ belonging to the different vertices V_λ of the tree T.

In the last three sections, the arrangement of the modular constituents of an

---

[3] Cf. R. Brauer, *On groups whose order contains a prime factor to the first order*, to appear soon.

[4] BN, §9. Cf. also R. Brauer and C. Nesbitt, *On the modular representations of groups of finite order*, University of Toronto Studies, Math. Series No. 4, 1937.

[5] For the definition of p-conjugate characters, cf. the list of notations at the end of this introduction.

[6] This improves the theorem that the degree of an irreducible representation divides the order of the group.

irreducible representation $\mathfrak{Z}_\lambda$ of a degree $z_\lambda \equiv 0 \pmod{p^{a-1}}$ is discussed. We assume here that the splitting field Ω is normal over the field of rational numbers and that the order of ramification e of p in Ω is equal to the number r_λ of p-conjugates; there exist fields Ω satisfying these conditions, but e can never be smaller than r_λ. It turns out that the arrangement of the modular constituents of $\mathfrak{Z}_\lambda$ is uniquely determined apart from a cyclic permutation of the constituents. If we restrict ourselves to the application of similarity transformations M whose coefficients are $\mathfrak{P}$-integers and whose determinant is prime to $\mathfrak{P}$, then the cyclic permutation must be the identity. The class of all representations $\mathfrak{Z}_\nu$ with $\mathfrak{P}$-integral coefficients in Ω, which are similar to $\mathfrak{Z}_\lambda$, splits into j subclasses of representations which are similar in this narrower sense. Here, j is the number of modular constituents of $\mathfrak{Z}_\lambda$, and each of the subclasses corresponds to one of the j cyclic permutations of the modular constituents.

NOTATION: $\mathfrak{G}$ is a group of order $g = p^a g'$ where p is a fixed prime and $(g', p) = 1$. The words "representation" and "character" always refer to representations in the field of all complex numbers and their characters, unless the word "modular" is added, in which case we mean a representation in a field of characteristic p. The distinct irreducible representations of $\mathfrak{G}$ are denoted by $\mathfrak{Z}_1, \mathfrak{Z}_2, \cdots, \mathfrak{Z}_k$, their characters by $\zeta_1, \zeta_2, \cdots, \zeta_k$ and their degrees by $z_1, z_2, \cdots, z_k$. Two characters ζ_κ and ζ_λ are p-conjugate if ζ_κ can be carried into ζ_λ by a change of the primitive p^ath roots of unity which leaves the g'th roots of unity unaltered. Then $\zeta_1, \zeta_2, \cdots, \zeta_k$ are distributed into "families" of p-conjugate characters; the number of members of the family of ζ_λ is usually denoted by r_λ. If G is an element of $\mathfrak{G}$, then $\zeta_\lambda(G)$ is the value of ζ_λ for this element G. The distinct, absolutely irreducible modular characters of G are denoted by $\varphi_1, \varphi_2, \cdots$. An element G of $\mathfrak{G}$ is p-regular if its order is prime to p. If its order is divisible by p, then G is p-singular.

1. Construction of a suitable splitting field[7]

Let $\mathfrak{G}$ be a group of finite order g, and consider an irreducible representation $\mathfrak{Z}$ of the group $\mathfrak{G}$ in the field of all complex numbers. We want to construct a splitting field Λ for $\mathfrak{Z}$ such that Λ is normal over the field P of rational numbers, and that the order of ramification e of a given prime p in Λ is as small as possible. Of course, Λ must contain the field $\mathrm{Z} = \mathrm{P}(\zeta)$ generated by the character ζ of $\mathfrak{Z}$, and this fact imposes a condition on e. The following lemma shows that, for suitable Λ, the number e has the smallest value which is compatible with this condition.

LEMMA 1: *There exist splitting fields of the form* $\mathrm{Z}(\tau)$ *where* τ *is a root of unity,* $\tau^s = 1$, *of an order* s *which is prime to* p.

PROOF: Let $\mathfrak{q}$ be a prime ideal of Z and $\mathfrak{Q}$ be a corresponding prime ideal of

[7] For the concepts of the theory of algebras used in §1, cf. M. Deuring, *Algebren*, Berlin 1935, and A. A. Albert, *Structure of algebras*, New York 1939.

$Z(\tau)$. According to a theorem of Hasse,[8] the field $Z(\tau)$ will be a splitting field, if for every $\mathfrak{q}$ the $\mathfrak{q}$-index $m_{\mathfrak{q}}$ of Z divides the $\mathfrak{Q}$-degree $n_{\mathfrak{q}}$ of $Z(\tau)$. Let $\mathfrak{q}$ be one of the prime ideals for which $m_{\mathfrak{q}} \neq 1$, and let r^{ν} be the highest power of the rational prime r which divides $m_{\mathfrak{q}}$. We shall show below that we can find a root of unity $\vartheta = \vartheta(\mathfrak{q}, r^{\nu})$ of an order $j = j(\mathfrak{q}, r^{\nu})$ prime to p, such that for every prime divisor $\mathfrak{Q}_{\vartheta}$ of q in $Z(\vartheta)$, the $\mathfrak{Q}$-degree is divisible by r^{ν}. If we then adjoin the numbers $\vartheta(\mathfrak{q}, r^{\nu})$ for all discriminant divisors $\mathfrak{q}$, and all prime powers r^{ν} dividing $m_{\mathfrak{q}}$, a field $Z(\tau)$ with the desired property is obtained.

The construction of ϑ is immediate for infinite prime ideals; any imaginary root of unity of an order prime to p can be taken. Let $\mathfrak{q}$ be finite and let q be the rational prime divisible by $\mathfrak{q}$. It will be sufficient to find a positive rational integer j which is prime to p, such that in a field of the j^{th} roots of unity either the order of ramification e_q or the degree f_q of the prime ideal divisors of q is divisible by a preassigned prime power r^t.[9] We distinguish several cases:

(a) if $r \neq p$, $r \neq q$ then assume that q belongs to the exponent ρ (mod r). We set $j = r^{\lambda}$ where λ is a positive rational integer, and have $q^{\rho j} \equiv 1 \pmod{j}$. The degree f_q equals the exponent to which q belongs (mod j). Hence $\rho \mid f_q$, but $f_q \mid \rho j$ and consequently, f_q is of the form ρr^{σ}. For sufficiently large λ, we shall have $\sigma \geqq t$, i.e. $r^t \mid f_q$.

(b) If $r = q$, $q \neq p$, take $j = r^{t+1}$. Then $e_q = r^t(r-1)$.

(c) If $r = p$, $q \not\equiv 1 \pmod{p}$, take $j = q^{r^t} - 1$. Then $(j, p) = 1, f_q = r^t$.

(d) If $r = p$, $q \equiv 1 \pmod{p}$, we may write $q^{r^{\lambda}} - 1 = c_{\lambda} r^{b_{\lambda}}$ where $(c_{\lambda}, r) = 1$, $b_{\lambda} > 0$.[10] Raising this equation to the r^{th} power we easily obtain $c_{\lambda-1} < c_{\lambda}$. For $j = c_{\lambda}$, $\lambda \geqq t$, we have $f_q = r^{\lambda} \geqq r^t$. In all these cases j is prime to p. This finishes the proof of the lemma.

Without restriction, we may assume that all g'^{th} roots of unity belong to $Z(\tau)$. We may further assume that τ is so chosen that it can be used simultaneously for all irreducible representations of $\mathfrak{G}$. If we set $K = P(\tau)$, we have

LEMMA 1*: *There exists a field* K *over the field of rational numbers with the following properties*:

(α) *The field* K *contains the* g'^{th} *roots of unity.*

(β) *The prime* p *is not ramified in* K: $(p) = \mathfrak{p}\mathfrak{m}$, *where* $\mathfrak{p}$ *is a prime ideal of* K, *and* $(\mathfrak{p}, \mathfrak{m}) = (1)$.

(γ) *Every irreducible representation* $\mathfrak{Z}$ *of* $\mathfrak{G}$ *can be written in the field* $K(\zeta)$ *obtained from* K *by adjoining the character* ζ *of* $\mathfrak{Z}$.

Two characters ζ and ζ' are algebraically conjugate with regard to K, if and only if they are p-conjugate. The degree r of $K(\zeta)$ with regard to K is, therefore equal to the number of characters in the family of ζ.

Let ϵ be a p^{α}th root of unity such that ζ lies in $K(\epsilon) = \Omega$; we may always take

[8] Cf. the books mentioned in footnote 7 or the original paper of Hasse, Math. Ann. **107**, p. 731 (1933).

[9] Since the field Z is fixed, the field $Z(\vartheta)$ will satisfy the condition above, if t is taken large enough.

[10] If $r = 2$, choose $\lambda \geqq 2$.

$1 \leqq \alpha \leqq a$. Let $\mathfrak{P}$ be the prime ideal divisor of $\mathfrak{p}$ in Ω.[11] The representation $\mathfrak{Z}$ can be written with $\mathfrak{P}$-integral coefficients belonging to Ω. When we replace every coefficient by its residue class (mod $\mathfrak{P}$), we obtain a modular representation $\bar{\mathfrak{Z}}$ whose coefficients belong to the field $\bar{\Omega}$ of residue classes of integers of Ω (mod $\mathfrak{P}$). After replacing $\mathfrak{Z}$ by a similar representation, we may assume that $\bar{\mathfrak{Z}}$ splits into irreducible constituents in $\bar{\Omega}$.

If $\bar{\mathfrak{F}}$ is any irreducible representation of $\mathfrak{G}$ in the algebraically closed extension field of $\bar{\Omega}$, then the traces of all the matrices are sums of modular g^{th} roots of unity. Since $\bar{\Omega}$ has the characteristic p, they are sums of g'^{th} roots of unity, and hence they belong to $\bar{\Omega}$. It follows that $\bar{\mathfrak{F}}$ can be written with coefficients in $\bar{\Omega}$.[12] Any modular representation which is irreducible in $\bar{\Omega}$ is absolutely irreducible. We may set

$$\mathfrak{Z} = \begin{pmatrix} \mathfrak{A}_{11} & \mathfrak{A}_{12} & \cdots \\ \mathfrak{A}_{21} & \mathfrak{A}_{22} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}, \tag{1}$$

where $\mathfrak{A}_{\kappa\lambda} \equiv 0 \pmod{\mathfrak{P}}$ for $\kappa < \lambda$, and where the $\mathfrak{A}_{\kappa\kappa}$ (mod $\mathfrak{P}$) are the absolutely irreducible modular constituents of $\mathfrak{Z}$.

2. On the number of p-conjugate characters

The degree of Ω over K is $m = \varphi(p^\alpha) = p^{\alpha-1}(p-1)$; we denote the conjugates of a number ω of Ω by $\omega, \omega', \cdots, \omega^{(m-1)}$, and use the corresponding notation for matrices and representations. Then $\mathfrak{Z}^{(\rho)}$ and $\mathfrak{Z}^{(\sigma)}$ are either non-similar or identical.

Let $\mathfrak{Z}(G)$ be the matrix representing the group element G, and denote by $\beta(G)$ the coefficient in the upper right corner of $\mathfrak{Z}(G)$ and by $\gamma(G)$ the coefficient in the lower left corner of $\mathfrak{Z}(G)$. The fundamental relations of I. Schur[13] for the coefficients of a representation give

$$\sum_{G \text{ in } \mathfrak{G}} \beta(G)^{(\rho)} \gamma(G^{-1})^{(\sigma)} = \eta_{\rho\sigma} g/z, \tag{2}$$

where z is the degree of $\mathfrak{Z}$, and $\eta_{\rho\sigma} = 0$ for $\zeta^{(\rho)} \neq \zeta^{(\sigma)}$, and $\eta_{\rho\sigma} = 1$ for $\zeta^{(\rho)} = \zeta^{(\sigma)}$. Let ω be any integer of Ω, and set

$$\xi_1(G) = \sum_{\rho=0}^{m-1} \omega^{(\rho)} \beta(G)^{(\rho)} = \operatorname{tr}(\omega\beta(G)); \qquad \xi_2(G) = \sum_{\sigma=0}^{m-1} \gamma(G)^{(\sigma)} = \operatorname{tr}(\gamma(G)), \tag{3}$$

where tr $(\cdots)$ denotes the trace of an element of Ω with regard to K. Then (2) yields

$$\sum_G \xi_1(G)\xi_2(G^{-1}) = (g/z) \cdot \sum_{\rho,\sigma} \eta_{\rho\sigma} \omega^{(\rho)}.$$

---

[11] We then have $\mathfrak{p} = \mathfrak{P}^e$, $e = p^{\alpha-1}(p-1)$.

[12] Cf. R. Brauer, Math. Zeitschr. 29, p. 79 (1929), p. 101. The fact used is equivalent to Wedderburn's theorem on finite division rings. The argument shows that $\bar{\mathfrak{F}}$ can be written in the field of the modular g'^{th} roots of unity and, therefore, in the field $\bar{\text{K}}$ of residue classes of integers of K (mod $\mathfrak{p}$).

[13] I. Schur, Sitzungsber. Preuss. Akad. d. Wiss. 1905 p. 406, theorem IV.

If r is the degree of $K(\zeta)$ over K, we have m/r characters $\zeta^{(\rho)}$ which are equal to a fixed $\zeta^{(\mu)}$. Hence

$$\sum_G \xi_1(G)\xi_2(G^{-1}) = \frac{gm}{zr}\operatorname{tr}(\omega). \tag{4}$$

The trace of any $\mathfrak{P}$-integer is divisible by $\mathfrak{p}^{\alpha-1}$.[14] This holds, in particular, for $\xi_1(G)$ and $\xi_2(G)$. For $\omega = 1$, we thus find

$$\frac{gm^2}{rz} \equiv 0 \pmod{\mathfrak{p}^{2\alpha-2}}.$$

Since the left side is rational, $\mathfrak{p}$ may be replaced by p (cf. lemma 1*, (β)). Since $m = \varphi(p^\alpha)$, this gives

Theorem 1: *If ζ is an irreducible character of degree z, and if r is the number of p-conjugate characters, then $g/(rz)$ is a p-integer for any prime p.*

We next assume that $\mathfrak{Z}$ is reducible (mod $\mathfrak{P}$). Then (1) shows that $\beta(G) \equiv 0$ (mod $\mathfrak{P}$). Set

$$v = p^{\alpha-1}, \qquad \Theta = (1-\epsilon^v)/(1-\epsilon) = 1 + \epsilon + \cdots + \epsilon^{v-1}. \tag{5}$$

Since ϵ^v is a primitive p^{th} root of unity, Θ is divisible by $\mathfrak{P}^{v-1}$. If a $\mathfrak{P}$-integer is divisible by $\mathfrak{P}^v$, its trace is divisible by $\mathfrak{p}^\alpha$.[15] For $\omega = \Theta$, we have $\omega\beta(G) \equiv 0$ (mod $\mathfrak{P}^v$) and, therefore, $\xi_1(G) \equiv 0 \pmod{\mathfrak{p}^\alpha}$. As before, $\xi_2(G) \equiv 0 \pmod{\mathfrak{p}^{\alpha-1}}$. From (5) it follows that $\operatorname{tr}(\Theta) = m$, and (4) now yields

$$gm^2/rz \equiv 0 \pmod{\mathfrak{p}^{2\alpha-1}}.$$

Hence

Lemma 2: *If ζ in theorem 1 is reducible as a modular character, then the p-integer g/rz is still divisible by p.*

3. Characters with a common modular constituent

We use a similar argument in order to study two irreducible representations $\mathfrak{Z}_1$ and $\mathfrak{Z}_2$ which have a modular constituent in common. We exclude the case when the characters ζ_1 and ζ_2 of $\mathfrak{Z}_1$ and $\mathfrak{Z}_2$ are p-conjugate. The $p^{\alpha\text{th}}$ root of unity ϵ may be so chosen that $\Omega = K(\epsilon)$ contains both ζ_1 and ζ_2; let $\mathfrak{H}$ be the Galois group of Ω with regard to K.

Each $\mathfrak{Z}_\mu$ can be written in the form (1), say $\mathfrak{Z}_\mu = (\mathfrak{A}^\mu_{\kappa\lambda})$. For at least one pair of indices i, j, we must have

$$\mathfrak{A}^1_{ii} \equiv \mathfrak{A}^2_{jj} \pmod{\mathfrak{P}}. \tag{6}$$

[14] Any $\mathfrak{P}$-integer λ can be written in the form $\lambda = \sum u_\mu\epsilon^\mu$ where $\mu = 0, 1, \cdots, m-1$ and the u_μ are $\mathfrak{p}$-integers of K. We have $\operatorname{tr}(\epsilon^\mu) = m$ if $\mu \equiv 0 \pmod{p^\alpha}$, $\operatorname{tr}(\epsilon^\mu) = -m/(p-1)$, if $\mu \equiv 0 \pmod{p^{\alpha-1}}$, $\mu \not\equiv 0 \pmod{p^\alpha}$, and $\operatorname{tr}(\epsilon^\mu) = 0$ in all other cases. This proves the statement.

[15] It is sufficient to prove this for numbers of the form $(1-\epsilon^v)\epsilon^\mu$, cf. footnote 14, and here it is evident.

Let $\gamma_1(G)$ be the coefficient in the upper left corner of $\mathfrak{A}^1_{ii}$, and $\gamma_2(G)$ the coefficient in the upper left corner of $\mathfrak{A}^2_{jj}$, and set $\beta(G) = \gamma_1(G) - \gamma_2(G)$ so that

$$\beta(G) = \gamma_1(G) - \gamma_2(G) \equiv 0 \pmod{\mathfrak{P}}. \tag{7}$$

Schur's relations[18] give here

$$\begin{gathered}\sum_G \gamma_1(G)^{(\rho)}\gamma_1(G^{-1})^{(\sigma)} = \frac{g}{z_1}\eta'_{\rho\sigma}; \qquad \sum_G \gamma_2(G)^{(\rho)}\gamma_2(G^{-1})^{(\sigma)} = \frac{g}{z_2}\eta''_{\rho\sigma};\\ \sum_G \gamma_1(G)^{(\rho)}\gamma_2(G^{-1})^{(\sigma)} = 0,\end{gathered} \tag{8}$$

where z_μ is the degree of $\mathfrak{Z}_\mu$, and $\eta^\mu_{\rho\sigma} = 0$ or 1 according as $\zeta_\mu^{(\rho)} \neq \zeta_\mu^{(\sigma)}$ or $\zeta_\mu^{(\rho)} = \zeta_\mu^{(\sigma)}$.

The same relations hold, when $\mathfrak{Z}_1 = \mathfrak{Z}_2$, and $\mathfrak{Z}_1$ contains one of its modular constituents more than once. We then have a formula (6) with $i \neq j$, and (7) and (8) are also true.

From (8) it follows that

$$\sum_G \beta(G)^{(\rho)}\beta(G^{-1})^{(\sigma)} = \frac{g}{z_1}\eta'_{\rho\sigma} + \frac{g}{z_2}\eta''_{\rho\sigma}. \tag{9}$$

We set (cf. (3))

$$\xi_1(G) = \operatorname{tr}(\omega\beta(G)), \qquad \xi_2(G) = \operatorname{tr}(\psi\beta(G)), \tag{10}$$

where ω and ψ are two integers of Ω. Then (9) gives

$$\sum_G \xi_1(G)\xi_2(G^{-1}) = \frac{g}{z_1}\sum_{\rho,\sigma}\eta'_{\rho\sigma}\omega^{(\rho)}\psi^{(\sigma)} + \frac{g}{z_2}\sum_{\rho,\sigma}\eta''_{\rho\sigma}\omega^{(\rho)}\psi^{(\sigma)}.$$

When $K(\zeta_\mu)$ corresponds to the subgroup $\mathfrak{L}_\mu$ of the Galois group $\mathfrak{H}$, this can be written in the form

$$\sum_G \xi_1(G)\xi_2(G^{-1}) = u_1\frac{g}{z_1} + u_2\frac{g}{z_2}, \tag{11}$$

$$u_\mu = \sum_{X_1 \equiv X_2 \pmod{\mathfrak{L}_\mu}} \omega^{X_1}\psi^{X_2}, \tag{12}$$

where in (12) the sum is to be extended over all pairs of elements X_1, X_2 of $\mathfrak{H}$, for which $X_1X_2^{-1}$ belongs to $\mathfrak{L}_\mu$.

We first choose $\omega = \Theta$, $\psi = 1$, where Θ is defined in (5). Then $\omega\beta(G) \equiv 0 \pmod{\mathfrak{P}^v}$, and as in §2, the trace $\xi_1(G)$ of $\omega\beta(G)$ is divisible by $\mathfrak{p}^\alpha$. Furthermore, $\xi_2(G)$ is divisible by $\mathfrak{p}^{\alpha-1}$, since it is a trace. If $\alpha = 1$, we may even state $\xi_2(G) \equiv 0 \pmod{\mathfrak{p}}$, because $\beta(G)$ is divisible by $\mathfrak{P}$. Finally, $u_\mu = \operatorname{tr}(\Theta)m/r_\mu = m^2/r_\mu$, since $\mathfrak{L}_\mu$ has the order m/r_μ and $\operatorname{tr}(\Theta) = m$. Then (11) yields

$$gm^2/r_1z_1 + gm^2/r_2z_2 \equiv 0 \pmod{\mathfrak{p}^{2\alpha-1}}.$$

For $\alpha = 1$, this congruence even holds $\pmod{\mathfrak{p}^2}$. If $(r_1, p) = 1$ and $(r_2, p) = 1$, we may choose $\alpha = 1$. As before, $\mathfrak{p}$ can be replaced by p. Since $m = p^{\alpha-1}(p-1)$, we have

LEMMA 3: *Let ζ_1 and ζ_2 be two characters of $\mathfrak{G}$ which are not p-conjugate but which have a modular constituent in common. If ζ_μ has r_μ p-conjugates, and if z_μ is the degree of ζ_μ, then*

$$g/r_1z_1 + g/r_2z_2 \equiv 0 \pmod{p}, \tag{13}$$

$$g/r_1z_1 + g/r_2z_2 \equiv 0 \pmod{p^2}, \quad \text{if} \quad (r_1, p) = 1 \quad \text{and} \quad (r_2, p) = 1. \tag{13*}$$

Secondly, we choose ω, ψ such that $\omega \equiv 0, \psi \equiv 0 \pmod{\mathfrak{P}^{v-1}}$. Then $\xi_1(G) \equiv \xi_2(G) \equiv 0 \pmod{p^a}$, and we obtain

LEMMA 4: *Suppose that ζ_1 and ζ_2 satisfy the assumptions of lemma 3. Let ω and ψ be two integers of $K(\epsilon)$ which are divisible by $\mathfrak{P}^{v-1}$, $v = p^{a-1}$. Then*

$$u_1g/z_1 + u_2g/z_2 \equiv 0 \pmod{p^{2a}}, \tag{14}$$

where u_μ is defined by (12).

According to a remark above, the same argument will hold, if $\mathfrak{Z}_1 = \mathfrak{Z}_2$ and $\mathfrak{Z}_1$ contains one of its modular constituents more than once.

LEMMA 5: *If ζ is a character which contains one of its modular constituents more than once, then for odd p,*

$$g/(rz) \equiv 0 \pmod{p^2}, \tag{15}$$

where z is the degree of ζ, and where the number r of p-conjugates of ζ is assumed to be prime to p.

4. On representations for which $rz \equiv 0 \pmod{p^a}$

We now prove

THEOREM 2: *Let ζ be an irreducible character of degree z which has r distinct p-conjugates. Then rz can be divisible by p^a only if z is divisible by p^a, i.e. when ζ is of the highest kind.*

PROOF: Assume that $rz \equiv 0 \pmod{p^a}$, $z \not\equiv 0 \pmod{p^a}$. Since the corresponding representation $\mathfrak{Z}$ is not of the highest kind, the block B of $\zeta = \zeta_1$ cannot consist of the family of ζ_1 only.[16] Hence we may find a character ζ_2 in B which is not p-conjugate to ζ_1, but has a modular constituent in common with ζ_1. Then (13) shows that also $r_2z_2 \equiv 0 \pmod{p^a}$. On the other hand, $z_2 \not\equiv 0 \pmod{p^a}$, because B is not of the highest kind. Consequently, ζ_2 satisfies the same assumptions as ζ_1. On considering chains of characters of B such that any two consecutive terms have a modular character in common, we conclude that every character ζ_ν of B satisfies the conditions $r_\nu z_\nu \equiv 0 \pmod{p^a}$, $z_\nu \not\equiv 0 \pmod{p^a}$, where z_ν again is the degree of ζ_ν and r_ν the number of p-conjugates.

It now follows from lemma 2 that all the ζ_ν of B are modular-irreducible, and this implies that they all are equal when considered as modular characters. To the block $B = B_\lambda$ there corresponds a part D_λ[17] of the matrix D of the decompo-

[16] R. Brauer, *On the connection between the ordinary and the modular characters of groups of finite order*, Ann. of Math., **42**, pp. 926–935, 1941, theorem 8.

[17] Cf. BN, §9, (28).

sition numbers of $\mathfrak{G}$. Our results show that the matrix D_λ consists of one column only, and all the coefficients are 1. All the ζ_ν have the same degree z, and if $z = p^\rho z'$ with $(z', p) = 1$, then all the r_ν are divisible by $p^{a-\rho}$. The number of rows in D_λ is equal to the sum of all r_ν belonging to the different families of B, and hence this number is larger than $p^{a-\rho}$.

There corresponds to B a part $C_\lambda = D_\lambda' D_\lambda$ of the matrix C of the Cartan invariants.[17] This C_λ is of degree 1; its only coefficient is equal to the number of rows of D_λ, and therefore is larger than $p^{a-\rho}$. On the other hand, the block B_λ is of type ρ, and then C_λ has $p^{a-\rho}$ as an elementary divisor,[18] which means that $C_\lambda = (p^{a-\rho})$. This gives a contradiction, and hence the theorem is proved.

5. Evaluation of the numbers u_μ in lemma 4

Assume that $\alpha \geqq 2$. Let $\boldsymbol{\rho}$ be a system of $p^{\alpha-1} = v$ rational integers ρ_1, $\rho_2, \cdots$ which form a complete residue system (mod $p^{\alpha-1}$), $\alpha \geqq 2$, and let $\boldsymbol{\sigma}$: $\sigma_1, \sigma_2, \cdots$ be a second system with the same properties. We set

$$\omega = \Sigma\, \epsilon^{\rho}, \qquad \psi = \Sigma\, \epsilon^{-\sigma},$$

where ρ ranges over the values of $\boldsymbol{\rho}$, and σ over those of $\boldsymbol{\sigma}$. If $\rho \equiv \tau \pmod{v}$, then $\epsilon^{\rho} \equiv \epsilon^{\tau} \pmod{\mathfrak{P}^v}$ when again $v = p^{\alpha-1}$. It follows that ω and ψ both are congruent to the number Θ in (5) and hence both are divisible by $\mathfrak{P}^{v-1}$ as assumed in lemma 4. Now (12) becomes

$$(16) \qquad u_\mu = \sum_{X_1 \equiv X_2 \pmod{\mathfrak{L}_\mu}} \sum_{\rho} \sum_{\sigma} \epsilon^{\rho X_1 - \sigma X_2} \qquad (X_1, X_2 \text{ in } \mathfrak{H}).^{19}$$

If ϵ^{λ} appears, all its conjugates appear with the same multiplicity. Consequently, if A_μ' terms of (16) are equal to 1, and B_μ' terms are primitive p^{th} roots of unity, we have $u_\mu = A_\mu' - B_\mu'/(p-1)$. The term on the right-hand side of (16) is 1, if $\rho X_1 \equiv \sigma X_2 \pmod{p^\alpha}$, i.e. if $\rho X_1 X_2^{-1} \equiv \sigma \pmod{p^\alpha}$. Similarly, the term is a primitive p^{th} root of unity, if this congruence holds (mod $p^{\alpha-1}$) but not (mod p^α). Now $X_1 X_2^{-1}$ is an element of $\mathfrak{L}_\mu$. Let A_μ denote the number of pairs (ρ, L), ρ in $\boldsymbol{\rho}$, L in $\mathfrak{L}_\mu$, for which ρL is congruent to one of the numbers $\sigma \pmod{p^\alpha}$, and B_μ the number of pairs, for which a congruence $\rho L \equiv \sigma$ (σ in $\boldsymbol{\sigma}$) holds (mod $p^{\alpha-1}$) but not (mod p^α). Then A_μ' and B_μ' are obtained from A_μ and B_μ by multiplying the latter by the order $\varphi(p^\alpha) = v(p-1)$ of $\mathfrak{H}$, and hence

$$(17) \qquad u_\mu = v(p-1)A_\mu - vB_\mu\,.$$

[18] In BN, §17, it was proved that at least one elementary divisor corresponding to a block of type α is divisible by $p^{a-\alpha}$. Actually, exactly one elementary divisor is equal to $p^{a-\alpha}$ while the other elementary divisors are smaller than $p^{a-\alpha}$. This can be proved by a generalization of the method of BN, §21; cf. below section 8 of this paper where the same method is used (for $p = 2$).

[19] We may understand ρX_1 and σX_2 as rational integers (mod p^α).

Since every ρL is congruent to one of the σ (mod $p^{\alpha-1}$), and since $\mathfrak{L}_\mu$ is of the order $l_\mu = v(p-1)/r_\mu$, we have $A_\mu + B_\mu = v^2(p-1)/r_\mu$. Then (17) becomes $u_\mu = vpA_\mu - v^3(p-1)/r_\mu$ and (14) takes the form

$$\frac{p^\alpha g}{z_1} A_1 + \frac{p^\alpha g}{z_2} A_2 - \left(\frac{p^{3\alpha-3}(p-1)g}{r_1 z_1} + \frac{p^{3\alpha-3}(p-1)g}{r_2 z_2}\right) \equiv 0 \pmod{p^{2\alpha}}.$$

According to (13), the last two terms are divisible by $p^{3\alpha-2}$, and since $3\alpha - 2 \geqq 2\alpha$, can be neglected. We then have

$$(18) \qquad gA_1/z_1 + gA_2/z_2 \equiv 0 \pmod{p^\alpha}.$$

If p is odd, we shall need the value of A_μ, only if r_μ is prime to p. The field $\mathrm{K}(\zeta_\mu)$ then is contained in the field $\mathrm{K}(\epsilon_p)$ where ϵ_p is a primitive p^{th} root of unity. If $\rho \not\equiv 0 \pmod v$, ($\rho$ in ϱ), then for every L in $\mathfrak{L}_\mu$ the number $\epsilon^{\rho L}$ is conjugate to ϵ^ρ with regard to $\mathrm{K}(\zeta_\mu)$. If ϵ^τ appears in the form $\epsilon^{\rho L}$, so do

$$\epsilon^\tau, \epsilon^\tau \epsilon_p, \cdots, \epsilon^\tau \epsilon_p^{p-1}$$

and each of them appears the same number of times. Exactly one of these quantities is of the form ϵ^σ, (σ in $\mathfrak{d}$). Hence, for a fixed ρ, one in every p of the elements L of $\mathfrak{L}_\mu$ satisfies the condition that ρL belongs to $\mathfrak{d}$. For $\rho \equiv 0 \pmod v$, the number of elements L of $\mathfrak{L}_\mu$, for which $\epsilon^{\rho L}$ has a fixed value, is divisible by $p^{\alpha-1}$. Hence, in the case $(r_\mu, p) = 1$, we have

$$(19) \qquad A_\mu \equiv l_\mu(v-1)/p \equiv v(p-1)(v-1)/r_\mu p \equiv p^{\alpha-2}/r_\mu \pmod{p^{\alpha-1}}.$$

The congruences (18) and (19) enable us to prove the following lemma:

LEMMA 6: *Let p be odd. If ζ_1 and ζ_2 satisfy the assumptions of lemma 3, and if $z_1 \equiv 0 \pmod{p^{\alpha-1}}$, then $r_2 \not\equiv 0 \pmod p$.*

PROOF: From theorem 2, it follows that $r_1 \not\equiv 0 \pmod p$. Assume that $r_2 \equiv 0 \pmod p$. If we choose for α the smallest value for which ζ_1 and ζ_2 belong to the field $\mathrm{K}(\epsilon)$ obtained from K by adjoining a $p^{\alpha\text{th}}$ root of unity ϵ, then r_2 must be divisible by $p^{\alpha-1}$. Theorem 2 shows that we have $z_2 \not\equiv 0 \pmod{p^{a-\alpha+1}}$ and, therefore, the second term in (18) is divisible by p^α. Since $(r_1, p) = 1$, we may use (19) for $\mu = 1$ and find that the first term in (18) is divisible by $p^{\alpha-1}$ only, which gives a contradiction. Hence $r_2 \not\equiv 0 \pmod p$ as was stated.

For $p = 2$, we choose $\alpha \geqq 3$. Here, r_μ is a power of 2. If $r_\mu = 2^j$, $1 \leqq j \leqq \alpha - 2$, the field $\mathrm{K}(\epsilon)$ has exactly three subfields of degree 2^j over K. One of them, say $\Gamma_j^{(0)}$, is obtained by adjoining a $2^{j+1\text{th}}$ root of unity to K. We denote the other two by $\Gamma_j^{(1)}$ and $\Gamma_j^{(2)}$, and set $\mathrm{K}(\epsilon) = \Gamma_{\alpha-1}^{(0)}$, $\mathrm{K} = \Gamma_0^{(1)}$. Then, by an argument similar to the one used for odd p, we may prove

$$(20) \qquad A_\mu \equiv 0 \pmod{2^{\alpha-1-j}}, \quad \text{if} \quad \mathrm{K}(\zeta_\mu) = \Gamma_j^{(0)}.$$

$$(21) \qquad A_\mu \equiv 2^{\alpha-2-j} \pmod{2^{\alpha-1-j}}, \quad \text{if} \quad \mathrm{K}(\zeta_\mu) = \Gamma_j^{(1)} \quad \text{or} \quad \Gamma_j^{(2)}.$$

LEMMA 7: *Let $p = 2$, and suppose that ζ_1 and ζ_2 satisfy the assumptions of lemma 3. If $r_1 z_1 \equiv 0 \pmod{2^{\alpha-1}}$ and if $\mathrm{K}(\zeta_1)$ is not of the form $\mathrm{K}(\bar{\epsilon})$ where $\bar{\epsilon}$ is a primitive*

$2^{j\text{th}}$ *root of unity with* $j \geq 2$, *then* $r_2z_2 \equiv 0 \pmod{2^{a-1}}$, *and* $K(\zeta_2)$ *is not of the form* $K(\bar{\epsilon})$.

Proof: From the assumptions, theorem 2, and (21), it follows that A_1g/z_1 is divisible by 2^{a-1}, but not by 2^a. According to (18), A_2g/z_2 then is also divisible by 2^{a-1} but not by 2^a. If we set $r_2 = 2^j$ and use (20), we see that $K(\zeta_2)$ cannot be one of the fields $\Gamma_j^{(0)}$, $(j = 1, 2, \cdots)$. If $K(\zeta_2) = \Gamma_j^{(1)}$ and $K(\zeta_2) = \Gamma_j^{(2)}$ we use (21) and conclude that z_2 is divisible by 2^{a-1-j}, which implies $r_2z_2 \equiv 0 \pmod{2^{a-1}}$.

6. Representations of a degree which is divisible by p^{a-1}

Let ζ be an irreducible character of degree $z \equiv 0 \pmod{p^{a-1}}$. If z is divisible by p^a, then ζ is of the highest kind,[20] and we will exclude this case in the following. We shall first assume that p is odd.[21] The block B to which $\zeta = \zeta_1$ belongs is not of the highest kind. Then B does not consist of the family of ζ only,[22] and we may therefore find a character ζ_2 which is not p-conjugate to $\zeta = \zeta_1$ but has a modular constituent in common with ζ_1 (cf. the analogous argument in §4). If z_μ is the degree of ζ_μ, and r_μ the number of p-conjugate characters, then theorem 2 and lemma 6 show that $r_1 \not\equiv 0$, $r_2 \not\equiv 0 \pmod{p}$. In (13*), lemma 3, the first term on the left side is divisible by p but not by p^2. The same must hold for the second term, and hence we must have $z_2 \equiv 0 \pmod{p^{a-1}}$, so that ζ_2 satisfies the same assumption as ζ. Continuing in this manner we see that the degree of every character of B is divisible by p^{a-1}. This gives

Theorem 3: *If the degree of one character of a block* B *is divisible by* p^{a-1}, *the same is true for all characters of* B.

In the notation of BN, the block B is of the type $a - 1$. If ζ_1 and ζ_2 are again two characters of such a block B which are not p-conjugate but have a modular constituent in common, then on multiplying (13*) by $r_1r_2z_1z_2/g \equiv 0 \pmod{p^{a-2}}$ the congruence $r_1z_1 + r_2z_2 \equiv 0 \pmod{p^a}$ is obtained. Now any two characters ζ_κ, ζ_μ can be joined by a chain $\zeta_\kappa, \zeta_\rho, \cdots, \zeta_\sigma, \zeta_\mu$ of characters, such that any two consecutive terms of the chain have a modular character in common without being p-conjugate. It follows that for any ζ_μ of B, we have $r_\mu z_\mu \equiv \pm r_1z_1 \pmod{p^a}$. Since $r_1z_1 \not\equiv 0 \pmod{p^a}$ and p is odd, for each μ only one of the signs can be used. Hence

Theorem 4: *The characters of a block* B *of type* $a - 1$ *can be distributed in two subsets* B′ *and* B″ *such that every character belongs to exactly one of these subsets. If* z_μ *is the degree of a character of* B, *and* r_μ *the number of p-conjugates, we may find a rational integer* N *such that* $r_\mu z_\mu \equiv N \pmod{p^a}$ *for all* ζ_μ *of* B′ *and* $r_\mu z_\mu \equiv -N \pmod{p^a}$ *for all* ζ_μ *in* B″. *If* ζ_μ *and* ζ_ν *are not p-conjugate and belong both to the same subset* $B^{(\tau)}$, *then they have no modular constituent in common.*

Finally, lemma 5 can be applied and yields

[20] Cf. BN, Part II.

[21] The case $p = 2$ will be treated in §8.

[22] Cf. footnote 16.

THEOREM 5: *If ζ belongs to a block* B *of type* $a - 1$, *then it contains each of its modular constituents only with the multiplicity* 1.

7. The matrix D_λ corresponding to a block B of type $a - 1$

NOTATION: $B = B_\lambda$ is a block of type $a - 1$ consisting of the ordinary characters $\zeta_1, \zeta_2, \cdots, \zeta_x$ and the modular characters $\varphi_1, \varphi_2, \cdots, \varphi_y$. The degree of ζ_μ is z_μ, the number of p-conjugates r_μ. There are w families of p-conjugate characters ζ_μ in B, and the ζ_μ are arranged so that $\zeta_1, \zeta_2, \cdots, \zeta_w$ lie in different families, while the characters of the subset B'_λ (theorem 4) come before the characters of the other subset B''_λ. For p-regular elements G of $\mathfrak{G}$, we have

$$\zeta_\mu(G) = \sum d_{\mu\nu}\varphi_\nu(G), \qquad (\mu = 1, 2, \cdots, x). \tag{22}$$

We set $D_\lambda = (d_{\mu\nu})$, $(\mu = 1, 2, \cdots, x; \nu = 1, 2, \cdots, y)$ and denote the matrix occupying the first w rows of D_λ by $\tilde{D}_\lambda$, i.e. $\tilde{D}_\lambda = (d_{\mu\nu})$, $(\mu = 1, 2, \cdots, w; \nu = 1, 2, \cdots, y)$. The matrix D_λ has the same rows as $\tilde{D}_\lambda$, the μ^{th} row of $\tilde{D}_\lambda$ appearing r_μ times in D_λ, $(\mu = 1, 2, \cdots, w)$.

The theorems 4 and 5 give at once

THEOREM 6: *Every column of $\tilde{D}_\lambda$ contains exactly two coefficients* 1, *one in a row corresponding to a character ζ_λ of B'_λ, and the other in a row corresponding to a character ζ_μ of B''_λ. All the other coefficients in the column are zero.*

We also can prove easily

THEOREM 7: *Let $\xi = (\xi_1, \xi_2, \cdots, \xi_w)$ be a row with w elements such that $\xi\tilde{D}_\lambda = 0$. Then ξ is a scalar multiple of the row $(\delta_1, \delta_2, \cdots, \delta_w)$, where $\delta_\mu = 1$ if ζ_μ belongs to B'_λ, and $\delta_\mu = -1$, if ζ_μ belongs to B''_λ.*

PROOF: Two characters $\zeta_1, \zeta_\mu (\mu = 1, 2, \cdots, w)$ of B can be joined by a chain of characters ζ_ρ, $(\rho = 1, 2, \cdots, w)$, such that two consecutive characters ζ_i, ζ_j of the chain have a modular character in common. Then ζ_i, ζ_j belong to different subsets B'_λ, B''_λ (theorem 4). There must be a column of $\tilde{D}_\lambda$ which contains the coefficients 1 in the i^{th} and the j^{th} row. Then $\xi\tilde{D}_\lambda = 0$ implies $\xi_i = -\xi_j$, and we obtain successively $\xi_\mu = \delta_\mu\xi_1$, $(\xi_1, \cdots, \xi_w) = \delta_1\xi_1(\delta_1, \cdots, \delta_w)$ which proves the theorem.

The rank of $\tilde{D}_\lambda$ is y.[23] Hence we have

COROLLARY 1: *The number w of families in* B *and the number y of modular characters of* B *are connected by the formula*

$$w = y + 1. \tag{23}$$

We further state

COROLLARY 2: *If $\tilde{\zeta}_\mu$ is the sum of the r_μ characters which are p-conjugate to $\zeta_\mu (\mu = 1, 2, \cdots, w)$, then $\delta_1\tilde{\zeta}_1(G) = \delta_\mu\tilde{\zeta}_\mu(G)$ for any p-singular element of G of $\mathfrak{G}$.*

Indeed, the numbers $\tilde{\zeta}_1(G), \tilde{\zeta}_2(G), \cdots, \tilde{\zeta}_w(G)$ form a solution of $\xi\tilde{D}_\lambda = 0$, (cf. equation (11) of the paper mentioned in footnote 16). Similarly, equation (6) of the same paper gives

---

[23] Cf. BN, §9.

COROLLARY 3: *Let* $d^i_{\mu\nu}$ *be the higher decomposition numbers,* $i > 0$, *and set* $\bar{d}^i_{\mu\nu} = \sum d^i_{\rho\nu}$ *where the sum extends over all the* r_μ *values* ρ *for which* ζ_ρ *is p-conjugate to* ζ_μ. *Then* $\delta_1 \bar{d}^i_{1\nu} = \delta_\mu \bar{d}^i_{\mu\nu}$.

For p-regular elements G, the formula (22), together with $(\delta_1, \cdots, \delta_w)\tilde{D}_\lambda = 0$ gives

COROLLARY 4: *For p-regular elements* G *of* $\mathfrak{G}$, *we have*

$$\sum_{\mu=1}^{w} \delta_\mu \zeta_\mu(G) = 0. \tag{24}$$

In particular, for $G = 1$, *this gives*

$$\sum_{\mu=1}^{w} \delta_\mu z_\mu = 0. \tag{25}$$

The matrix $\tilde{D}_\lambda$ has w minors of degree $y = w - 1$, and these, if properly arranged and taken with suitable signs, form a solution of $\xi\tilde{D}_\lambda = 0$. Let Δ be a fixed minor. Then, according to theorem 7, every minor has the value $\pm\Delta$. The minor obtained by removing the μ^{th} column of $\tilde{D}_\lambda$ will appear $r_1 r_2 \cdots r_w/r_\mu$ times as a minor of D_λ. We now form the determinant of $C_\lambda = D'_\lambda D_\lambda$. On the one hand, this determinant has the value p.[24] On the other hand, its value is the sum of the squares of all the minors of D_λ. This gives

$$\sum_{\mu=1}^{w} \Delta^2(r_1 r_2 \cdots r_w/r_\mu) = p.$$

Consequently, we must have $\Delta = \pm 1$, and we find

COROLLARY 5: *The numbers* r_μ *satisfy the equation*

$$r_1 r_2 \cdots r_w\left(\frac{1}{r_1} + \frac{1}{r_2} + \cdots + \frac{1}{r_w}\right) = p. \tag{26}$$

The numbers r_μ are divisors of $p - 1$; (26) shows that any two of them are relatively prime.

We associate w distinct points $P_1, P_2, \cdots, P_w$ with the characters $\zeta_1, \zeta_2, \cdots, \zeta_w$; we join P_i and P_j when ζ_i and ζ_j have a modular character in common. The linear graph thus obtained characterizes the matrix $\tilde{D}_\lambda$ completely (apart from the arrangements of the columns). We prove

COROLLARY 6: *The linear graph associated with a block* B *of type* $a - 1$ *is a tree.*

PROOF: This follows from the facts that the graph is connected and has one more vertex than it has edges.

Each point P_μ which lies only on one of the edges corresponds to a row of $\tilde{D}_\lambda$ which contains only one coefficient 1, and this means that ζ_μ is a modular-irreducible character.

---

[24] The determinant of C is a power of p, cf. R. Brauer, *On the Cartan invariants of groups of finite order*, Ann. of Math. **42**, p. 53, 1941. It follows from the result of footnote 18, that the exponent must be 1.

COROLLARY 7: *Every block* B *of type* $a - 1$ *contains at least two characters* ζ_μ *which are not p-conjugate and which are modular-irreducible characters.*

8. The case $p = 2$

For $p = 2$, several of the preceding proofs do not hold. We shall now treat this case, using a different kind of argument.

Let ζ be an irreducible character of degree z where $z \equiv 0 \pmod{2^{a-1}}$ but $z \not\equiv 0 \pmod{2^a}$; here a is again the highest exponent for which $2^a = p^a$ divides g. From theorem 2, it follows that ζ is p-conjugate only to itself, i.e. $r = 1$, and ζ belongs to K. Since K is not of the form $K(\bar{\epsilon})$ where $\bar{\epsilon}$ is a primitive $2^{j\text{th}}$ root of unity with $j \geqq 2$, we may apply lemma 7, §5. It follows that if the character ζ_μ has a modular constituent in common with ζ_1 then $r_\mu z_\mu \equiv 0 \pmod{2^{a-1}}$, (where z_μ denotes again the degree of ζ_μ, and r_μ the number of 2-conjugates). Further, $K(\zeta_\mu)$ also is not of the form $K(\bar{\epsilon})$. Therefore, lemma 7 again can be applied to ζ_μ and any character which has a modular constituent in common with ζ_μ without being p-conjugate to ζ_μ. We finally see that the congruence $r_\mu z_\mu \equiv 0 \pmod{2^{a-1}}$ is true for any character ζ_μ of the block B to which ζ belongs.

Let $\zeta_1, \zeta_2, \cdots, \zeta_x$ be the ordinary characters of B, and $\varphi_1, \varphi_2, \cdots, \varphi_y$ the modular characters of B, and set

$$z_\mu = 2^{\tau_\mu} z'_\mu, \qquad (z'_\mu, 2) = 1. \tag{27}$$

It now follows easily that

$$r_\mu = 2^{a-1-\tau_\mu}. \tag{28}$$

Suppose that $\tau_1 = m$ is the smallest of the numbers $\tau_1, \tau_2, \cdots, \tau_x$.

We consider the sums

$$S_{\mu\nu} = \sum_{\kappa=1}^{k_0} g_\kappa \zeta_\mu(G_\kappa)\zeta_\nu(G_\kappa^{-1}), \tag{29}$$

where G_κ ranges over a system of representatives for the 2-regular classes of $\mathfrak{G}$, and where g_κ denotes the number of elements in the class of G_κ. The value of $S_{\mu\nu}$ is a rational integer. Since $g_\kappa \zeta_\mu(G_\kappa)/z_\mu = \omega_{\mu\kappa}$ is an algebraic integer, it follows easily that $S_{\mu\nu}$ is divisible by z_μ and hence by 2^m. On the other hand, we have[25] $\omega_{\mu\kappa} \equiv \omega_{1\kappa}$ modulo a prime ideal divisor of 2. Hence

$$S_{\mu\nu} = z_\mu \sum_\kappa \omega_{\mu\kappa}\zeta_\nu(G_\kappa^{-1}) \equiv z_\mu \sum_\kappa \omega_{1\kappa}\zeta_\nu(G_\kappa^{-1}) = \frac{z_\mu}{z_1} S_{1\nu} \pmod{2^{\tau_\mu+1}}.$$

On applying the same argument to $S_{1\nu} = S_{\nu 1}$, we arrive at

$$S_{\mu\nu} \equiv 0 \pmod{2^m}, \tag{30}$$

$$S_{\mu\nu} = S_{\nu\mu} \equiv \frac{z_\mu z_\nu}{z_1^2} S_{11} \pmod{2^{\tau_\mu+1}}. \tag{31}$$

[25] BN, §9.

If the block B is the λ^{th} of the blocks $B_1, B_2, \cdots$ of $\mathfrak{G}$, and if $D_\lambda = (d_{\mu\nu})$, $C_\lambda = (c_{\mu\nu})$ are the parts[26] of the matrices D and C corresponding to the block B_λ, then (22) gives for the matrix $(S_{\mu\nu})_{\mu,\nu}$

$$(S_{\mu\nu})_{\mu,\nu} = \left(\sum_{\kappa} \sum_{\rho,\sigma=1}^{y} g_\kappa d_{\mu\rho} \varphi_\rho(G_\kappa)\, d_{\nu\sigma} \varphi_\sigma(G_\kappa^{-1})\right)_{\mu,\nu} = D_\lambda(gC_\lambda^{-1})D'_\lambda.\,^{27}$$

The relation (31) can be written in the form

$$(S_{\mu\nu})_{\mu,\nu} = D_\lambda(gC_\lambda^{-1})D'_\lambda = S_{11}\left(\frac{z_\mu z_\nu}{z_1^2}\right)_{\mu,\nu} + H, \tag{32}$$

where H is a matrix in which every coefficient in both the μ^{th} row and μ^{th} column is divisible by $2^{\tau_\mu+1}$. The first matrix on the right side in (32) has the rank 1; every coefficient in both the μ^{th} row and μ^{th} column is divisible by 2^{τ_μ}.

We now proceed to discuss the elementary divisors of $(S_{\mu\nu})$, choosing the ring of all 2-integers as the underlying domain. If C_λ has the elementary divisors $e_1, e_2, \cdots, e_y$ with $e_\nu = 2^{\epsilon_\nu}$, then

$$|\, C_\lambda \,| = e_1 e_2 \cdots e_y = 2^{\epsilon_1+\epsilon_2+\cdots+\epsilon_y}.\,^{28}$$

The elementary divisors of gC_λ^{-1} are $2^{a-\epsilon_y}, \cdots, 2^{a-\epsilon_2}, 2^{a-\epsilon_1}$, and, since the columns of D_λ are linearly independent (mod 2),[29] the elementary divisors of $D_\lambda(gC^{-1})D'_\lambda$ are given by

$$2^{a-\epsilon_y}, 2^{a-\epsilon_{y-1}}, \cdots, 2^{a-\epsilon_1}, \qquad 0, \cdots, 0. \tag{33}$$

On the other hand, the right side of (32) can also be used for a discussion of these elementary divisors. If M is a minor of degree j of $(S_{\mu\nu})$ involving the rows $\mu_1, \mu_2, \cdots, \mu_j$, then a simple computation shows that M is divisible by 2 to the exponent $\sum \tau_\mu + j - 1$, $(\mu = \mu_1, \mu_2, \cdots, \mu_j)$ because of the properties of the matrices on the right side of (32). If the j characters $\zeta_\mu(G)$, $(\mu = \mu_1, \mu_2, \cdots, \mu_j)$, are linearly dependent for 2-regular elements G of $\mathfrak{G}$, then $M = 0$, as follows from (29).

Let us now choose a maximal system of characters of B which are linearly independent for 2-regular elements $\mathfrak{G}$. Such a system consists of y characters. We first take as many characters as possible with $\tau_\mu = m$, then as many as possible with $\tau_\mu = m + 1$, etc. We may assume that the characters chosen are $\zeta_1, \cdots, \zeta_y$. Let β_ρ be the number of characters $\zeta_1, \cdots, \zeta_y$ for which $\tau_\mu = m + \rho - 1$, $(\rho = 1, 2, \cdots, s)$, such that $\beta_1 > 0, \beta_2 \geq 0, \cdots, \beta_{s-1} \geq 0, \beta_s > 0$. Then

$$\begin{gathered} \tau_1 = m, \qquad \tau_y = m + s - 1; \\ m \leq \tau_\mu \leq m + s - 1 \quad \text{for} \quad \mu = 1, 2, \cdots, y. \end{gathered} \tag{34}$$

---

[26] BN, §9, (28).
[27] Cf. BN, (21).
[28] The determinant is actually a power of 2, cf. footnote 24.
[29] BN, §19.

Of course, we have

$$\beta_1 + \beta_2 + \cdots + \beta_s = y.$$

It now follows that every minor of degree y of $(S_{\mu\nu})$ is divisible by 2^l with

$$l = \beta_1 m + \beta_2(m + 1) + \cdots + \beta_s(m + s - 1) + \beta_1 + \beta_2 + \cdots + \beta_s - 1,$$

$$l = my - 1 + \sum_{\sigma=1}^{s} \sigma\beta_\sigma.$$

Similarly, the minors of degree $y - 1$ are divisible by $2^{l'}$ with

$$l' = \beta_1 m + \beta_2(m + 1) + \cdots + \beta_{s-1}(m + s - 2) \\ + (\beta_s - 1)(m + s - 1) + \beta_1 + \cdots + \beta_s - 2,$$

$$l' = m(y - 1) - 1 - s + \sum_{\sigma=1}^{s} \sigma\beta_\sigma.$$

On comparing this with (33), we find

$$ya - (\epsilon_1 + \epsilon_2 + \cdots + \epsilon_y) \geqq my - 1 + \sum_{\sigma=1}^{s} \sigma\beta_\sigma,$$

$$(y - 1)a - (\epsilon_2 + \epsilon_3 + \cdots + \epsilon_y) \geqq (y - 1)m - 1 - s + \sum_{\sigma=1}^{s} \sigma\beta_\sigma.$$ [30]

From (28) we see that $r_1 r_2 \cdots r_y$ is the power of 2 with the exponent

$$(a - 1 - m)\beta_1 + (a - 2 - m)\beta_2 + \cdots + (a - s - m)\beta_s = (a - m)y - \sum_{\sigma=1}^{s} \sigma\beta_\sigma.$$

Therefore, the two inequalities can be written in the form

$$| C_\lambda | = e_1 e_2 \cdots e_y \leqq 2 r_1 r_2 \cdots r_y, \tag{35}$$

$$e_2 e_3 \cdots e_y \leqq 2^{m-a+s+1} r_1 r_2 \cdots r_y. \tag{36}$$

No two of the characters $\zeta_1, \zeta_2, \cdots, \zeta_y$ can be 2-conjugate. Letting w be the number of families of 2-conjugate characters which belong to the block $B = B_\lambda$, we have $w > y$.[31] We arrange $\zeta_{y+1}, \cdots, \zeta_x$ in such a manner that $\zeta_1, \zeta_2, \cdots, \zeta_w$ all lie in different families. Denote the μ^{th} row of D_λ by $\mathfrak{d}_\mu$, and let

$$D_\lambda = \begin{pmatrix} \mathfrak{d}_1 \\ \vdots \\ \mathfrak{d}_x \end{pmatrix} = \begin{pmatrix} \bar{D}_\lambda \\ \mathfrak{d}_{w+1} \\ \vdots \\ \mathfrak{d}_x \end{pmatrix} = \begin{pmatrix} T \\ \mathfrak{d}_{y+1} \\ \vdots \\ \mathfrak{d}_x \end{pmatrix},$$

---

[30] This holds in the case $y = 1$ also; we have here $s = 1$, $\beta_1 = 1$, and $\epsilon_2 + \cdots + \epsilon_y$ is to be taken equal to 0, and $e_2 \cdots e_y$ equal to 1.

[31] Cf. footnote 16.

so that $\tilde{D}_\lambda$ contains the first w rows of D_λ, and T the first y rows. Then D_λ has the same rows as $\tilde{D}_\lambda$, the μ^{th} row of $\tilde{D}_\lambda$ appearing r_μ times in D_λ. Let Δ_1, $\Delta_2, \cdots$ be the minors of degree y of D_λ; the determinant $| T |$ appears $r_1 r_2 \cdots r_y$ times among these minors. Consequently, we have

$$(37) \qquad | C_\lambda | = | D'_\lambda D_\lambda | = \sum \Delta_\kappa^2 \geqq r_1 r_2 \cdots r_y | T |^2.$$

Here, $| T | \neq 0$, since $\zeta_1(G), \cdots, \zeta_y(G)$ are linearly independent even if G ranges over the 2-regular elements only. If all the minors of degree y of $\tilde{D}_\lambda$ except $| T |$ vanish, then $\mathfrak{d}_{y+1} = 0$, which is impossible. Hence the inequality sign must hold in (37). On comparing (35) and (37), we find

$$(38) \qquad | T | = \pm 1,$$

and, since every r_ν and e_ν is a power of 2, we have

$$(39) \qquad | C_\lambda | = 2 r_1 r_2 \cdots r_y$$

Then (36) yields

$$(40) \qquad e_1 \geqq 2^{a-m-s}.$$

Because of (38), we may set

$$\mathfrak{d}_\nu = h_{\nu 1}\mathfrak{d}_1 + \cdots + h_{\nu y}\mathfrak{d}_y,$$

where the $h_{\nu\rho}$ are rational integers. Then $D_\lambda = HT$ with $H = (h_{\nu\rho})$. It follows that $C_\lambda = T'H'HT$, and this shows that $H'H$ has the same elementary divisors as C_λ. In particular, the coefficient in the y^{th} row, y^{th} column of $H'H$ must be divisible by e_1, and hence, by (40), we have

$$(41) \qquad \sum_{\nu=1}^{x} h_{\nu y}^2 \equiv 0 \pmod{2^{a-m-s}}.$$

When ζ_ν is 2-conjugate to ζ_y, then $\mathfrak{d}_\nu = \mathfrak{d}_y$ and $h_{\nu y} = 1$. There must exist characters ζ_ν which are not 2-conjugate to ζ_y and for which $h_{\nu y} \neq 0$.[32] Using (34), we obtain

$$\sum_{\nu=1}^{x} h_{\nu y}^2 > r_y = 2^{a-1-\tau_y} = 2^{a-m-s}$$

and, therefore, (41) yields

$$(42) \qquad \sum_{\nu=1}^{x} h_{\nu y}^2 \geqq 2^{a+1-m-s} = 2r_y.$$

Consider now the minor Δ_κ consisting of the rows $1, 2, \cdots, y-1, \nu$. Its value is $h_{\nu y} | T | = \pm h_{\nu y}$. If ζ_ν is 2-conjugate to one of the characters $\zeta_1, \cdots, \zeta_{y-1}$,

[32] In §5 of the paper mentioned in footnote 16, it is shown that it is impossible to find a linear combination of the columns of $\tilde{D}_\lambda$, such that exactly one row contains a term $\neq 0$. This implies the statement.

then the minor vanishes. It is easily seen that there are $r_1 r_2 \cdots r_{y-1}$ minors of D_λ with the value $\pm h_{\nu y}$; and (37), (39) and (42) imply that

$$(43)\quad 2r_1 r_2 \cdots r_y = |C_\lambda| = \sum \Delta_\kappa^2 \geqq r_1 r_2 \cdots r_{y-1} \sum_{\nu=1}^{x} h_{\nu y}^2 \geqq 2r_1 \cdots r_{y-1} r_y .$$

Since actually the equality sign holds, it follows that those minors, which have not been taken into account, must all vanish. If $y > 1$, it follows that the minors formed by means of the rows 2, 3, $\cdots$, y, ν must vanish unless ζ_ν is 2-conjugate to ζ_1. The value of such a minor is $\pm h_{\nu 1}$, and we obtain a contradiction, since there must exist characters ζ_ν, which are not 2-conjugate to ζ_1, for which $h_{\nu 1} \neq 0$.[32]

It follows that $y = 1$, i.e. that D_λ has only one column. The coefficient in the first row must be 1, because of (38). We set $h_{\nu y} = h_\nu$,

$$D_\lambda = \begin{pmatrix} 1 \\ h_2 \\ \vdots \\ h_x \end{pmatrix}.$$

The relation (43), (28), and (34) then imply

$$(44)\qquad 1 + h_2^2 + \cdots + h_x^2 = 2r_y = 2r_1 = 2^{a-m}.$$

Since ζ_1 has degree $z_1 = 2^m z'$, the degree $z_2 = 2^{\tau_2} z_2'$ of ζ_2 must be equal to $h_2 z_1$, which shows that h_2 is divisible by $2^{\tau_2 - m}$. The left side of (44) is at least equal to $r_1 + r_2 h_2^2$, since r_1 terms 1 and r_2 terms h_2 appear. Because of (28) and (34) we have

$$2^{a-m} \geqq r_1 + r_2 h_2^2 \geqq 2^{a-1-m} + 2^{a-1-\tau_2} 2^{2\tau_2 - 2m} \geqq 2^{a-m}.$$

We readily see that $w = 2$, $\tau_2 = m$, $r_1 = r_2 = 2^{a-1-m}$, $h_2 = 1$. The block B_λ consists of two families. All of its characters have the same degree $z_1 = 2^m z_1'$.[33] Since it was assumed that B contains a character of degree $2^{a-1}z'$, $((z', 2) = 1)$, it follows that $m = a - 1$ and hence $r_1 = r_2 = 1$,

$$D_\lambda = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

This proves the theorems 3, 4, 5, 6, and 7, and the corollaries of §7, for $p = 2$.

9. Applications

THEOREM 8: *If $\mathfrak{G}$ is a group of order $g = p^a q^b r^c$, (p, q, r distinct primes) with $a \leqq 2$, then $\mathfrak{G}$ possesses irreducible representations $\mathfrak{Z}$ besides the 1-representation [1] whose degree z is a power of q, and also irreducible representations $\mathfrak{Z}_1 \neq$ [1] whose degree z_1 is a power of r.*

PROOF: Consider the p-block B of representations which contains the repre-

[33] All the facts derived so far hold for any block B which contains a character ζ_ν with the following properties: (1) ζ_ν is not of the highest kind. (2) $r_\nu z_\nu \equiv 0 \pmod{2^{a-1}}$. (3) $K(\zeta_\nu)$ cannot be obtained from K by adjoining a $2^{j\text{th}}$ root of unity with $j \geqq 2$.

sentation [1]. This B is not of the highest kind, and if $a = 2$, it is not of the type $a - 1 = 1$. It follows from theorem 3 that the degrees of all representations of B are prime to p. We apply the relation[34] $\sum \zeta_\mu(R)\zeta_\mu(S) = 0$ where μ ranges over the indices belonging to B, and where R is p-regular and S is p-singular. For $R = 1$, we have $\zeta_\mu(R) = z_\mu$. The term corresponding to the character [1] is 1. At least one other term $\zeta_\mu(1)\zeta_\mu(S) = z_\mu\zeta_\mu(S)$ must be prime to r, and then z_μ must be a power of q. In the same manner we see that a character $\zeta_\nu \neq [1]$ has a degree z_ν which is a power of r.

If $\mathfrak{G}$ is equal to its commutator group $\mathfrak{G}'$, then z_μ and z_ν cannot be 1 since [1] is the only linear representation of $\mathfrak{G}$. In particular, this will be so, if G is simple.

Assume that $r^c = 4$ or $r^c = 3$. Then $\mathfrak{G}$ must have a representation of one of the degrees 2, 3, or 4. Since all linear groups of these degrees are known,[35] we obtain easily

THEOREM 9: *The only simple group of an order* $4p^aq^b$, *(p, q primes) with* $a \leqq 2$ *is the alternating group* $\mathfrak{A}_5$ *of order* 60. *The only simple groups of an order* $3p^aq^b$, *(p, q primes) with* $a \leqq 2$ *are the groups* $A_5 \cong LF(2, 5)$ *of order* 60 *and the group* $LF(2, 7)$ *of order* 168.

10. On $\mathfrak{P}$-similiar representations

Let $\mathfrak{Z}$ be an irreducible representation of $\mathfrak{G}$ with $\mathfrak{P}$-integral coefficients in an algebraic number field Ω, where $\mathfrak{P}$ is a prime ideal divisor of p. If $\mathfrak{Z}_1$ is a second representation with $\mathfrak{P}$-integral coefficients in the same field Ω, it may happen that $\mathfrak{Z}$ and $\mathfrak{Z}_1$ are similar, but that the corresponding modular representations $\bar{\mathfrak{Z}}$ and $\bar{\mathfrak{Z}}_1$ are not similar. Indeed, if

$$P^{-1}\mathfrak{Z}P = \mathfrak{Z}_1, \tag{45}$$

we may assume that the coefficients of P are $\mathfrak{P}$-integers which are not all divisible by $\mathfrak{P}$. Going over to residue classes mod $\mathfrak{P}$ (which again will be indicated by a bar) we obtain

$$\bar{\mathfrak{Z}}\bar{P} = \bar{P}\bar{\mathfrak{Z}}_1$$

and $\bar{P} \neq 0$, so that $\bar{\mathfrak{Z}}$ and $\bar{\mathfrak{Z}}_1$ are intertwined. However, this proves $\bar{\mathfrak{Z}} \sim \bar{\mathfrak{Z}}_1$ only if the determinant of P is not divisible by $\mathfrak{P}$. We shall say that $\mathfrak{Z}$ and $\mathfrak{Z}_1$ are $\mathfrak{P}$-similar, if P in (45) can be chosen in accordance with these conditions, i.e. with $\mathfrak{P}$-integral coefficients and a determinant $\not\equiv 0 \pmod{\mathfrak{P}}$. The class of all representations $\mathfrak{Z}_1$ which are similar to $\mathfrak{Z}$ and have $\mathfrak{P}$-integral coefficients thus breaks up into subclasses of $\mathfrak{P}$-similar representations.

Conversely, if $\mathfrak{T}$ is any representation of $\mathfrak{G}$ in the field $\bar{\Omega}$ of residue classes (mod $\mathfrak{P}$), which is similar to $\bar{\mathfrak{Z}}$, then we may find a representation $\mathfrak{Z}_1$, which is $\mathfrak{P}$-similar to $\mathfrak{Z}$, such that $\bar{\mathfrak{Z}}_1 = \mathfrak{T}$.

[34] R. Brauer and C. Nesbitt, University of Toronto Studies, Math. Ser. No. 4, 1937, theorem VIII.

[35] H. F. Blichfeldt, *Finite collineation groups*, Chicago 1917.

In the general case of (45), we may assume that P appears in the normal form of the theory of elementary divisors (with regard to the domain of all $\mathfrak{P}$-integers) if we replace $\mathfrak{Z}$ and $\mathfrak{Z}_1$ by $\mathfrak{P}$-similar representations. We thus may set

$$(46)\qquad P = \begin{pmatrix} 0 & 0 & \cdots & \pi^{s-1} I_{m_s} \\ \multicolumn{4}{c}{\dotfill} \\ 0 & \pi I_{m_2} & \cdots & 0 \\ I_{m_1} & 0 & \cdots & 0 \end{pmatrix} = (\pi^{s-\kappa} \delta_{\kappa+\lambda, s+1} I_{m_\lambda})_{\kappa,\lambda},$$

where $m_1 > 0,\ m_2 \geqq 0, \cdots, m_{s-1} \geqq 0,\ m_s > 0$ are rational integers, I_m is the unit matrix of degree m, and π is a $\mathfrak{P}$-integer which satisfies $\pi \equiv 0 \pmod{\mathfrak{P}}$, $\pi \not\equiv 0 \pmod{\mathfrak{P}^2}$. If we set

$$(47)\qquad \mathfrak{Z} = \begin{pmatrix} \mathfrak{A}_{11} & \mathfrak{A}_{12} & \cdots & \mathfrak{A}_{1s} \\ \mathfrak{A}_{21} & \mathfrak{A}_{22} & \cdots & \mathfrak{A}_{2s} \\ \multicolumn{4}{c}{\dotfill} \\ \mathfrak{A}_{s1} & \mathfrak{A}_{s2} & \cdots & \mathfrak{A}_{ss} \end{pmatrix}, \qquad \mathfrak{Z}_1 = \begin{pmatrix} \mathfrak{B}_{11} & \mathfrak{B}_{12} & \cdots & \mathfrak{B}_{1s} \\ \mathfrak{B}_{21} & \mathfrak{B}_{22} & \cdots & \mathfrak{B}_{2s} \\ \multicolumn{4}{c}{\dotfill} \\ \mathfrak{B}_{s1} & \mathfrak{B}_{s2} & \cdots & \mathfrak{B}_{ss} \end{pmatrix},$$

where $\mathfrak{A}_{\kappa\lambda}$ has $m_{s-\kappa+1}$ rows and $m_{s-\lambda+1}$ columns, and $\mathfrak{B}_{\kappa\lambda}$ has m_κ rows and m_λ columns, then (45) gives

$$(48)\qquad \pi^{\lambda-1} \mathfrak{A}_{s+1-\kappa, s+1-\lambda} = \pi^{\kappa-1} \mathfrak{B}_{\kappa\lambda}.$$

This shows that in (47), all the terms above the main diagonal in $\mathfrak{Z}$ and $\mathfrak{Z}_1$ are congruent to 0 (mod $\mathfrak{P}$). In the modular sense, $\mathfrak{Z}$ and $\mathfrak{Z}_1$ split into the (reducible or irreducible) constituents

$$\bar{\mathfrak{A}}_{11} = \bar{\mathfrak{B}}_{ss},\ \bar{\mathfrak{A}}_{22} = \bar{\mathfrak{B}}_{s-1,s-1},\ \cdots,\ \bar{\mathfrak{A}}_{ss} = \bar{\mathfrak{B}}_{11}.$$

If $s = 1$, then $\mathfrak{Z}$ and $\mathfrak{Z}_1$ are $\mathfrak{P}$-similar. We certainly have this case when $\mathfrak{Z}$ is modular-irreducible in $\bar{\Omega}$.

THEOREM 10: *If the representation $\mathfrak{Z}$ remains irreducible as a modular representation, then every representation $\mathfrak{Z}_1$ with $\mathfrak{P}$-integral coefficients, which is similar to $\mathfrak{Z}$, is $\mathfrak{P}$-similar to $\mathfrak{Z}$.*

Remark: This theorem can always be applied if the degree z of $\mathfrak{Z}$ is divisible by p^a.

Further, if it is known that no fixed coefficient of a representation $\mathfrak{P}$-similar to $\mathfrak{Z}$ is divisible by $\mathfrak{P}^l$ for every G in $\mathfrak{G}$, then (48) shows that $s \leqq l$.

11. On the arrangement of the modular constituents of an irreducible representation $\mathfrak{Z}$ of type $a - 1$

We shall say that the algebraic splitting field Ω of the representation $\mathfrak{Z}$ is a *normal splitting field of least ramification*, if Ω is normal over the field of rational numbers P, and if the order of ramification of the fixed rational prime p is the same for Ω as for the subfield P(ζ) obtained by adjoining the character ζ of $\mathfrak{Z}$ to P. The existence of such fields Ω follows from lemma 1. Denote by K the field of inertia of the prime ideal divisor $\mathfrak{P}$ of p. Then

$$\mathfrak{P}^r = \mathfrak{p}, \quad (p) = \mathfrak{p}\mathfrak{q}, \quad \text{with} \quad (\mathfrak{p}, \mathfrak{q}) = (1),$$

where $\mathfrak{p}$ is a prime ideal divisor of p in K, and where r is the number of characters which are p-conjugate to ζ. As is easily seen, the degree of $K(\zeta)$ over K is r, and hence $K(\zeta) = \Omega$.

We now prove

LEMMA 8: *Let $\mathfrak{Z}$ be an irreducible representation of degree z with $z \equiv 0 \pmod{p^{a-1}}$ such that all the coefficients $a_{\kappa\lambda}(G)$ of the matrix $\mathfrak{Z}(G)$ representing G lie in the normal splitting field Ω of least ramification. For each pair (κ, λ) there exists a group element G_0 such that $a_{\kappa\lambda}(G_0)$ and $a_{\lambda\kappa}(G_0)$ are not both divisible by $\mathfrak{P}$. For each (κ, λ), there exists a group element G_0^* such that either $a_{\kappa\lambda}(G_0^*)$ is not a $\mathfrak{P}$-integer or $a_{\lambda\kappa}(G_0^*) \not\equiv 0 \pmod{\mathfrak{P}^2}$.*

PROOF: (a) Assume that $a_{\kappa\lambda}(G) \equiv a_{\lambda\kappa}(G) \equiv 0 \pmod{\mathfrak{P}}$ for every G in $\mathfrak{G}$. We now apply the method of §2 setting $\beta(G) = a_{\kappa\lambda}(G)$, $\gamma(G) = a_{\lambda\kappa}(G)$, $\omega = 1$. We have here $(r, p) = 1$ because of theorem 2, and hence $r \mid p - 1$, $m = \varphi(p)$. Then $\xi_1(G) \equiv 0$, $\xi_2(G) \equiv 0 \pmod{\mathfrak{p}}$, and (4) gives a contradiction.

(b) If $a_{\kappa\lambda}(G)$ is a $\mathfrak{P}$-integer and $a_{\lambda\kappa}(G) \equiv 0 \pmod{\mathfrak{P}^2}$, we multiply the κ^{th} row by an element $\pi \equiv 0 \pmod{\mathfrak{P}}$, for which $\pi \not\equiv 0 \pmod{\mathfrak{P}^2}$, and divide the κ^{th} column by π. The similar representation thus obtained satisfies the assumption of the part (a) of this proof. Therefore, we again obtain a contradiction.

Suppose now that the coefficients of $\mathfrak{Z}$ are $\mathfrak{P}$-integers. It follows at once from lemma 8 that for the corresponding modular representation $\bar{\mathfrak{Z}}$ all the Loewy constituents[36] are irreducible. This implies that the arrangement of the irreducible (modular) constituents of $\mathfrak{Z}$ is uniquely determined.

Further, if $\mathfrak{Z}$ and $\mathfrak{Z}_1$ both have $\mathfrak{P}$-integral coefficients and are similar, then it follows from lemma 8 that $s \leqq 2$ in the notation of §10, (46), (47). If $s = 1$, $\mathfrak{Z}$ and $\mathfrak{Z}_1$ are $\mathfrak{P}$-similar. If $s = 2$, we have by (48),

$$(49)\quad \mathfrak{Z} = \begin{pmatrix} \mathfrak{A}_{11} & \pi\mathfrak{B}_{21} \\ \mathfrak{A}_{21} & \mathfrak{A}_{22} \end{pmatrix}, \qquad P = \begin{pmatrix} 0 & \pi I_{m_2} \\ I_{m_1} & 0 \end{pmatrix}, \qquad \mathfrak{Z}_1 = \begin{pmatrix} \mathfrak{A}_{22} & \pi\mathfrak{A}_{21} \\ \mathfrak{B}_{21} & \mathfrak{A}_{11} \end{pmatrix}.$$

The modular-irreducible constituents of $\mathfrak{Z}_1$ are, of course, the same as those of $\mathfrak{Z}$, but we see that the arrangement in which they appear is a cyclic permutation H of the arrangement in $\mathfrak{Z}$; $H \neq 1$. Since all the modular constituents of $\mathfrak{Z}$ are distinct (theorem 5), it follows that $\mathfrak{Z}$ and $\mathfrak{Z}_1$ are not $\mathfrak{P}$-similar.

If the representation $\mathfrak{Z}$ with $\mathfrak{P}$-integral coefficients breaks up into two modular constituents, then we may set $\mathfrak{Z}$ in the form given by the first equation (49). If we define P, $\mathfrak{Z}_1$ by the other equations (49), then $\mathfrak{Z}_1$ is similar to $\mathfrak{Z}$. This shows that any cyclic permutation of the modular constituents of $\mathfrak{Z}$ can be effected by a transition to a similar representation $\mathfrak{Z}_1$.

THEOREM 11: *Let $\mathfrak{Z}$ be an irreducible representation of a degree $z \equiv 0 \pmod{p^{a-1}}$ whose coefficients are $\mathfrak{P}$-integers of a normal splitting field Ω of least ramification. Let j be the number of modular-irreducible constituents of $\mathfrak{Z}$. The class of all representations of $\mathfrak{G}$, whose coefficients are $\mathfrak{P}$-integers of Ω and which are similar to $\mathfrak{Z}$, splits into j subclasses of $\mathfrak{P}$-similar representations. In all representations*

[36] Cf. e.g. R. Brauer, Trans. Amer. Math. Soc. 49, p. 502, 1941.

of a fixed subclass, the modular constituents appear in the same arrangement. The j possible arrangements are obtained from the arrangement in $\mathfrak{Z}$ by the j cyclic permutations.

It is clear that lemma 8 will, in general, not remain valid if Ω is replaced by an extension field. Then theorem 11 also will not hold.

Let Ω be the normal splitting field of least ramification mentioned in lemma 1; τ is a root of unity of an order prime to p. The splitting group of the prime ideal $\mathfrak{P}$ contains the substitution T which transforms τ into τ^p and leaves the p^{a} th roots of unity fixed. Then T transforms any (ordinary or modular) character χ into a conjugate character χ^T. It is easily seen that if an irreducible character ζ splits into the modular characters $\varphi_\alpha, \varphi_\beta, \cdots, \varphi_\rho$ in this arrangement, then ζ^{T^κ} will split into the modular characters $\varphi_\alpha^{T^\kappa}, \varphi_\beta^{T^\kappa}, \cdots, \varphi_\rho^{T^\kappa}$. Assume again that the degree z of ζ is divisible by p^{a-1} where p is odd. If one of the constituents φ of ζ is left invariant by T^κ,then ζ and ζ^{T^κ} have a modular constituent φ in common. Since they have the same degree, they must be p-conjugate (cf. theorem 4[37]). It then follows easily that all the modular constituents of ζ will admit the substitution T^κ. The same will hold for all the characters ζ_μ which have a modular constituent in common with ζ, and finally for all the ζ_μ of the block B of ζ. Hence

THEOREM 12: *Let* B *be a block of type $a - 1$, and denote by T the substitution which replaces the g'^{th} roots of unity by their p^{th} powers but leaves the $p^{a\text{th}}$ roots of unity invariant. If a power T^κ of T transforms one of the modular characters of* B *into itself, it transforms every modular character of* B *into itself; every ordinary character ζ of* B *is transformed into a p-conjugate character by T^κ (p odd).*

12. Real characters of type $a - 1$

To every representation $\mathfrak{Z}$, there belongs a contragredient representation $\mathfrak{Z}^*$ of the same degree z; the characters ζ and ζ^* of $\mathfrak{Z}$ and $\mathfrak{Z}^*$ are conjugate complex, $\zeta^* = \bar{\zeta}$. If z is divisible by p^{a-1}, p odd,[38] then either ζ and ζ^* are p-conjugate, or they have no modular constituent in common, as follows from theorem 4.[37] Consequently, if ζ contains a modular character φ with $\varphi^* = \varphi$, then ζ and ζ^* must be p-conjugate. It is easily seen that, if the modular constituents of $\mathfrak{Z}$ are

$$\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_j \tag{49}$$

in this arrangement, those of $\mathfrak{Z}^*$ are

$$\mathfrak{F}_j^*, \mathfrak{F}_{j-1}^*, \cdots, \mathfrak{F}_1^*. \tag{50}$$

On the other hand, using a suitable splitting field Ω, we easily find that in any representation p-conjugate to $\mathfrak{Z}$, the modular constituents appear in the same arrangement as in $\mathfrak{Z}$. It now follows from theorem 11 that if ζ (G) is real for p-regular G the constituents (49) and (50) must be the same apart from a

[37] The number of p-conjugate characters is the same for both characters.

[38] The following theorems 13 and 14 are trivial for $p = 2$; cf. §8.

cyclic permutation. If $\mathfrak{F}_1 \sim \mathfrak{F}_\rho^*$, then $\mathfrak{F}_\nu \sim \mathfrak{F}_{\rho+1-\nu}^*$ where we set $\mathfrak{F}_{j+\mu} = \mathfrak{F}_\mu$. For odd j, there will be exactly one value of ν for which $\nu \equiv \rho + 1 - \nu \pmod{j}$, i.e. $\mathfrak{F}_\nu \sim \mathfrak{F}_\nu^*$. For even j, we have either two values ν or no value ν. Hence

THEOREM 13: *Let $\mathfrak{Z}$ be a representation of degree $z \equiv 0 \pmod{p^{a-1}}$ with the character ζ. If ζ is not real for p-regular elements, none of the modular constituents of ζ is real. If ζ is real for p-regular elements and contains j modular constituents, then, for odd j, exactly one of these modular constituents is real; if j is even, either two or none of them are real.*

If, in particular, all the modular characters of the block are real, then each ζ contains one or two modular constituents. In the tree corresponding to the block B of ζ, each vertex lies on at most two sides. Hence

THEOREM 14: *If all the modular characters of a block* B *of type $a - 1$ are real, then the corresponding tree is an open polygon.*

UNIVERSITY OF TORONTO
TORONTO, CANADA

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

ON THE DENSITY OF THE SUM OF SETS OF POSITIVE INTEGERS. II[1]

BY ALFRED BRAUER

In memory of I. Schur

(Received February 28, 1941)

Introduction

This paper is the continuation of my paper in the Annals of Mathematics vol. 39 (1938), pp. 322–340. Originally this continuation was to appear immediately after publication of the first part and was to contain improvements of some theorems of I. Schur by means of results of the first part. The publishing was postponed again and again, in order to improve the results. In this work I also obtained improvements of some theorems of my previous paper.

In 1, I state the results of the first part which I use in this paper.[2] In the following, let again $A_1, A_2, \cdots, A_n$ be sets of positive integers with the positive densities $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$, and $\gamma < 1$ be the density of the sum $A_1 + A_2 + \cdots + A_n$. Improving some results of the first part (see below 1, Theorems V, X, and XI) I obtain instead of V the following

THEOREM: *Suppose* $\alpha_n \geqq 1 - n\alpha_1$, *then we have*

$$\gamma \geqq \frac{n-1}{n} + \frac{\alpha_n}{n}.$$

Suppose $\alpha_1 \geqq 1 - n\alpha_2$, *then we have*

$$\gamma \geqq \frac{n-1}{n} + \frac{\alpha_1}{n}.$$

If we denote $\alpha_1 + \alpha_2 + \cdots + \alpha_n$ by σ_n, we obtain instead of XI:

$$(1) \qquad \begin{cases} \gamma \geqq \frac{9}{10}\sigma_2; \qquad \gamma \geqq \frac{81}{92}\sigma_3 > .8804\sigma_3; \qquad \gamma \geqq \frac{972}{1117}\sigma_4 > .8701\sigma_4; \\ \gamma \geqq \frac{34020}{39197}\sigma_n > .8679\sigma_n \qquad \text{for } n \geqq 5. \end{cases}$$

Moreover I obtain instead of X:

$$\gamma \geqq \alpha_1 + \tfrac{5}{8}\alpha_2.$$

If the conjecture is correct that always

$$(2) \qquad \gamma \geqq \alpha_1 + \alpha_2 + \cdots + \alpha_n,$$

[1] Presented to the American Mathematical Society, December 27, 1939.

[2] An interesting survey on the problems connected with this paper was given by H. Rohrbach in his paper *Einige neuere Untersuchungen über die Dichte in der additiven Zahlentheorie*, Jahresbericht der Deutschen Mathematiker-Vereinigung v. 48 (1938), pp. 199–236.

then one would obtain immediately for every n and m with $n \geqq m$ that

$$\gamma \geqq \frac{n}{m}(\alpha_1 + \alpha_2 + \cdots + \alpha_m). \tag{3}$$

In particular, it follows from (2) that the sum of the densities satisfies the inequality

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n < 1 \tag{4}$$

for $\gamma < 1$, and more generally

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m < m/n. \tag{5}$$

Even the special cases (4) and (5) are still unproved. One only knows that (4) is true for $n = 2$ and (5) moreover for $m = 1$. The former follows from the theorem of Schnirelmann, the latter from the theorem of Khintchine (see below, Theorems I and II).

In the following, I prove that (5) is true at all events if n and m satisfy the condition

$$\frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{n-m+1} \leqq \frac{m}{n}. \tag{6}$$

For $m = 2$, the condition (6) is always satisfied; for $m = 3$ it is satisfied if $n \geqq 5$, and for fixed m certainly if

$$n \geqq \tfrac{1}{2}(m^2 + m - 2).$$

Moreover I prove that (5) holds also for $m = 3$ and $n = 4$.

It follows from the theorem of Khintchine that the sum of n sets $A_1 + A_2 + \cdots + A_n$ is the set of all positive integers if $\alpha_1 \geqq 1/n$. It now follows that the same is true if

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m \geqq m/n$$

where m and n satisfy the condition (6), or where $m = 3$, $n = 4$.

I. Schur[3] proved

$$\sigma_3 < 1.3191; \qquad \sigma_4 < 1.3913; \qquad \sigma_5 < 1.4360;$$

$$\sigma_n < 1.6263 \qquad \text{for } n > 5,$$

in approximation to the conjecture (4). I improve this result as follows:

$$\sigma_3 < 21/20 = 1.05; \qquad \sigma_4 < 97/90 < 1.0778; \qquad \sigma_5 < 43/39 < 1.1026;$$

$$\sigma_n < \frac{39197n}{34020n + 5177} < 1.1522 \qquad \text{for } n > 5.$$

[3] I. Schur, Über den Begriff der Dichte in der additiven Zahlentheorie, Sitzungsberichte der Preussischen Akademie der Wissenschaften Phys.-Math. Klasse 1936, pp. 269–297.

In order to approximate (2) and (3) for every n and m with $m \leqq n$ we shall determine constants $c_n^{(m)}$ as large as possible such that

$$\gamma \geqq c_n^{(m)}(\alpha_1 + \alpha_2 + \cdots + \alpha_m).$$

For the special case $n = m$ we obtain such constants from (1). I. Schur[4] proved that one can choose

$$(7) \qquad c_n^{(1)} = n \quad \text{and} \quad c_n^{(m)} = \frac{nc_{n-1}^{(m-1)}}{c_{n-1}^{(m-1)} + (n^2 - n)^{\frac{1}{2}}} \quad \text{for } m > 1.$$

In the following I shall show that we can choose $c_n^{(1)} = n$ and for $m > 1$

$$(8) \qquad \begin{cases} c_n^{(m)} = \dfrac{n(n+2)c_{n-1}^{(m-1)}}{(n+2)c_{n-1}^{(m-1)} + n^2 + n - 1} & \text{if } c_{n-1}^{(m-1)} > \dfrac{n+1}{n+2}, \\[2ex] c_n^{(m)} = c_{n-1}^{(m-1)} & \text{if } c_{n-1}^{(m-1)} \leqq \dfrac{n+1}{n+2}. \end{cases}$$

It will be proved that this gives a better estimate than (7) for $m \geqq 2$.

For $n > m$ we set $n - m = d$; for a fixed value of d we denote by z the greatest integer such that

$$\log (z - 1) + 1/(z + 1) \leqq d + \log d + 2.$$

It then follows from (8) that

$$(9) \qquad \begin{cases} c_n^{(m)} > \dfrac{n+1}{m + \log n - \log d} & \text{for } n \leqq z, \\[2ex] c_n^{(m)} > \dfrac{z+1}{z+2} > \dfrac{e^{d+1}d + 1}{e^{d+1}d + 2} & \text{for } n \geqq z. \end{cases}$$

in approximation to (3).

Let m be fixed and n sufficiently large. I. Schur proved

$$c_n^{(m)} > n/m - 1/4$$

whereas I obtain the better estimate

$$c_n^{(m)} > n/m - \epsilon$$

from (9) for every $\epsilon > 0$.

1. Summary of former results

Definition 1: *Let $A_1, A_2, \cdots, A_n$ be sets of positive integers. The sum $S = A_1 + A_2 + \cdots + A_n$ is the set of all different positive integers representable in the form*

$$e_1a_1 + e_2a_2 + \cdots + e_na_n \quad \text{with} \quad a_\nu \text{ in } A_\nu; \qquad e_\nu = 0, 1 \quad \text{for} \quad \nu = 1, 2, \cdots, n.$$

[4] L. c., footnote 2.

DEFINITION 2: *Let $A(x)$ be the number of elements of A less than x. The density α of A is the lower bound of $A(x)/x$ for $x = 1, 2, \cdots$.*

THEOREM I (Schnirelmann): *If $n = 2$ and $\alpha_1 + \alpha_2 \geqq 1$, then we have $\gamma = 1$.*

In the following we assume $\gamma < 1$.

THEOREM II (Khintchine): $\gamma \geqq n\alpha_1$.

THEOREM III (Khintchine): *Suppose $0 \leqq \mu < \alpha < 1/n$ and for $x = 1, 2, \cdots$*

$$A_\nu(x) \geqq \alpha x - \frac{\nu - 1}{n} \qquad (\nu = 1, 2, \cdots n - 1),$$

$$A_n(x) \geqq (1 - n\alpha)x - n\mu.$$

Denote by S the sum $A_1 + A_2 + \cdots + A_n$, then

$$S(x) \geqq (1 - \alpha)x - \mu.$$

THEOREM IV (Schur): $$\gamma \geqq \alpha_2/(1 - \alpha_1) \qquad (n = 2).$$

THEOREM V [1][5]: *If $\alpha_1 < 1/n$ and $\alpha_n \geqq 1 - n\alpha_1$, then we have*

$$\gamma \geqq 1 - \alpha_1.$$

If $\alpha_2 < 1/n$ and $\alpha_1 \geqq 1 - n\alpha_2$, then we have

$$\gamma \geqq 1 - \alpha_2.$$

THEOREM VI [6]: $$\gamma \geqq \alpha_\nu + \sum_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} \frac{1}{1 + [1/\alpha_\mu]}.$$

THEOREM VII [14]: *Let $k \geqq 3$ be any integer and $k/(k-1)^2 \geqq \alpha_1$ then we have for $n = 2$*

$$\gamma \geqq \alpha_2 + \frac{k-1}{k}\alpha_1.$$

THEOREM VIII [15] and [16]: *Denote by β the density of $A_2 + A_3 + \cdots + A_n$. Then we have*

$$\gamma \geqq \beta + \frac{n+1}{n+2}\alpha_1 \geqq \alpha_n + \frac{3}{4}\alpha_{n-1} + \frac{4}{5}\alpha_{n-2} + \cdots + \frac{n+1}{n+2}\alpha_1.$$

THEOREM IX [17]: *For every a with $0 < a < 1$ we have*

$$\gamma \geqq a\beta + \alpha_1\left(n - \frac{n^2 + n - 1}{n + 2}a\right).$$

THEOREM X [17a]: $$\gamma \geqq \alpha_1 + \tfrac{4}{5}\alpha_2.$$

THEOREM XI [18]: $$\gamma \geqq \frac{8}{9}(\alpha_1 + \alpha_2); \qquad \gamma \geqq \frac{120}{139}(\alpha_1 + \alpha_2 + \alpha_3);$$

$$\gamma \geqq \frac{2880}{3361}(\alpha_1 + \alpha_2 + \cdots + \alpha_n) \qquad \text{for } n \geqq 4.$$

[5] The number in parentheses refers to the numeration in the first part of this paper, Annals of Mathematics v. 39 (1938), pp. 322–340.

In what follows these theorems will be cited by I, II, and so on.

2. Proof of the conjecture (5) for small values of m

First we prove

THEOREM 1: *Let n and m satisfy the conditions $1 < m \leqq n$ and*

$$\frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{n-m+1} \leqq \frac{m}{n}. \tag{10}$$

Then we have for $\gamma < 1$

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m < \frac{m}{n}. \tag{11}$$

PROOF: For $m = n$ the condition (10) is satisfied only for $m = n = 2$. In this case, (11) follows from I. Therefore we may now assume that $m < n$.

Let r be any integer with $0 \leqq r \leqq m$. Denote by β_r the density of $A_1 + A_2 + \cdots + A_r$ and by β_r^* the density of $A_{r+1} + A_{r+2} + \cdots + A_n$; we set $\beta_0 = 0$. Then it follows from II that

$$\beta_r \geqq r\alpha_1 \quad \text{and} \quad \beta_r^* \geqq (n - r)\alpha_{r+1}. \tag{12}$$

Because of I and $\gamma < 1$ we have $\beta_r + \beta_r^* < 1$, and it follows from (12) that

$$r\alpha_1 + (n - r)\alpha_{r+1} < 1,$$

$$\frac{r\alpha_1}{n-r} + \alpha_{r+1} < \frac{1}{n-r} \tag{13}$$

We add the inequalities (13) for $r = 0, 1, \cdots, m - 1$ after multiplying the first of them by a non-negative number w which will be determined later. Then we obtain

$$\begin{aligned} \alpha_1\left(w + \frac{1}{n-1} + \frac{2}{n-2} + \cdots + \frac{m-1}{n-m+1}\right) + \alpha_2 + \alpha_3 + \cdots + \alpha_m \\ < \frac{w}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{n-m+1}. \end{aligned} \tag{14}$$

On the other hand, it follows from (10) that

$$\frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{n-m+1} \leqq m,$$

$$\frac{n}{n-1} - 1 + \frac{n}{n-2} - 1 + - \cdots + \frac{n}{n-m+1} - 1 \leqq 1,$$

$$\frac{1}{n-1} + \frac{2}{n-2} + \cdots + \frac{m-1}{n-m+1} \leqq 1. \tag{15}$$

Therefore, setting

$$w = 1 - \frac{1}{n-1} - \frac{2}{n-2} - \cdots - \frac{m-1}{n-m+1}$$

we have $w \geqq 0$, hence

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m < \frac{1}{n}\left(1 - \frac{1}{n-1} - \frac{2}{n-2} \cdots - \frac{m-1}{n-m+1}\right) + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{n-m+1} = \frac{m}{n}$$

because of (14), and the theorem is proved.

It is easy to examine the condition (10) for small values of m. For $m = 2$ it is always satisfied, for $m = 3$ if $n \geqq 5$, and so on. Since (10) is equivalent to (15) and since

$$\frac{1}{n-1} + \frac{2}{n-2} + \cdots + \frac{m-1}{n-m+1} < \frac{1 + 2 + \cdots + m - 1}{n-m+1} = \frac{m(m-1)}{2(n-m+1)}$$

the inequality (11) is proved if

$$\frac{m(m-1)}{2(n-m+1)} \leqq 1,$$

$$n \geqq \tfrac{1}{2}m(m-1) + m - 1 = \tfrac{1}{2}(m^2 + m - 2).$$

COROLLARY 1: *If* (10) *holds and* (11) *is not satisfied, then the sum is the set of all positive integers.*

COROLLARY 2: *Let* m, k *and* t *be integers,* $m \leqq t$ *and* $n = kt$; *if*

$$\frac{1}{t-1} + \frac{1}{t-2} + \cdots + \frac{1}{t-m+1} \leqq \frac{m}{t},$$

then we have

$$\alpha_1 + \alpha_{k+1} + \alpha_{2k+1} + \cdots + \alpha_{(m-1)k+1} < \frac{m}{n}.$$

PROOF: Let $B_{\mu+1}$ be the sum $A_{\mu k+1} + A_{\mu k+2} + \cdots + A_{(\mu+1)k}$ for $\mu = 0, 1, \cdots, t-1$ and $\beta_{\mu+1}$ the density of $B_{\mu+1}$. It follows from II that

$$\beta_{\mu+1} \geqq k\alpha_{\mu k+1}. \tag{16}$$

Applying Theorem 1 to the sets $B_1, B_2, \cdots, B_t$ we have

$$\beta_1 + \beta_2 + \cdots + \beta_m < \frac{m}{t},$$

therefore, because of (16),

$$\alpha_1 + \alpha_{k+1} + \cdots + \alpha_{(m-1)k+1} < \frac{m}{kt} = \frac{m}{n}.$$

In particular, it follows for $m = 2$ and $n = 2k$ that

$$\alpha_1 + \alpha_{k+1} < \frac{2}{n}.$$

For odd n we infer

COROLLARY 3: *For $n = 2k + 1$ and $m = 2$ we have*

$$\alpha_1 + \alpha_{k+1} < \frac{2}{n}.$$

PROOF: We have

$$k\alpha_1 + (k + 1)\alpha_{k+1} < 1,$$

hence

$$(k + \tfrac{1}{2})(\alpha_1 + \alpha_{k+1}) < 1$$

because of $\alpha_1 \leqq \alpha_{k+1}$. This proves Corollary 3.

COROLLARY 4: *For $m = 3$ and $n = 4$ we have*

$$\alpha_1 + \alpha_2 + \alpha_3 < \tfrac{3}{4}.$$

PROOF: Suppose $\alpha_1 + \alpha_2 + \alpha_3 \geqq \frac{3}{4}$. Then we must have $\alpha_2 > \frac{1}{4}$ since $\alpha_1 + \alpha_3 < \frac{1}{2}$ by Corollary 2. It now follows from VI that the density α_{12} of $A_1 + A_2$ satisfies the condition

$$\alpha_{12} \geqq \alpha_1 + \tfrac{1}{4}.$$

Therefore we have by I and II

$$\begin{aligned} \alpha_{12} &+ 2\alpha_3 < 1, \\ \alpha_1 &+ 2\alpha_3 < 1 - \tfrac{1}{4}, \\ \alpha_1 &+ \alpha_2 + \alpha_3 < \tfrac{3}{4} \end{aligned}$$

since $\alpha_2 \leqq \alpha_3$. This gives a contradiction, and the corollary is proved.

3. Generalisation of Theorem VIII

THEOREM 2: *Let $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$ be the densities of the sets $A_1, A_2, \cdots, A_n$ and $\gamma < 1$ be the density of $A_1 + A_2 + \cdots + A_n$. Let $\alpha_{k-l,k}$ be the density of $A_{k-l} + A_{k-l+1} + \cdots + A_k$ for $n \geqq k > l \geqq 1$. Then we have*

$$\alpha_{k-l,k} \geqq \alpha_k + \frac{n-k+3}{n-k+4}\alpha_{k-1} + \frac{n-k+4}{n-k+5}\alpha_{k-2} + \cdots + \frac{n-k+l+2}{n-k+l+3}\alpha_{k-l}.$$

For $k = n$ and $l = n - 1$ we obtain Theorem VIII.

PROOF: In the first place we consider the case $l = 1$; we have to prove

$$\alpha_{k-1,k} \geqq \alpha_k + \frac{n-k+3}{n-k+4}\alpha_{k-1}.$$

If

$$\alpha_{k-1} \leqq (n-k+4)/(n-k+3)^2,$$

the statement follows from VII. Suppose now

$$\alpha_{k-1} > (n-k+4)/(n-k+3)^2.$$

Then we have

$$\frac{(n-k+3)^2}{n-k+4}\,\alpha_{k-1} > 1, \tag{17}$$

It follows from I and II that

$$(n-k+1)\alpha_k + \alpha_{k-1} < 1. \tag{18}$$

By (17) and (18) we have

$$\alpha_k + \frac{n-k+3}{n-k+4}\,\alpha_{k-1} < \frac{1}{n-k+1} - \frac{\alpha_{k-1}}{n-k+1} + \frac{n-k+3}{n-k+4}\,\alpha_{k-1}$$

$$< \frac{(n-k+3)^2}{(n-k+1)(n-k+4)}\,\alpha_{k-1} - \frac{\alpha_{k-1}}{n-k+1} + \frac{(n-k+1)(n-k+3)}{(n-k+1)(n-k+4)}\,\alpha_{k-1}$$

$$= \frac{(n-k+3)(2n-2k+4)}{(n-k+1)(n-k+4)}\,\alpha_{k-1} - \frac{\alpha_{k-1}}{n-k+1}$$

$$= 2\alpha_{k-1}\left\{1 + \frac{2}{(n-k+1)(n-k+4)}\right\} - \frac{\alpha_{k-1}}{n-k+1} \leqq 2\alpha_{k-1} \leqq \alpha_{k-1,k}$$

because of II.

Let us assume now that the theorem is proved for every k and for $l = 1, 2, \cdots, \lambda - 1$ with $\lambda < k$. Then we have

$$\alpha_{k-\lambda+1,k} \geqq \alpha_k + \frac{n-k+3}{n-k+4}\,\alpha_{k-1} + \cdots + \frac{n-k+\lambda+1}{n-k+\lambda+2}\,\alpha_{k-\lambda+1} \tag{19}$$

We have to prove the theorem for $l = \lambda$. If

$$\alpha_{k-\lambda} \leqq (n-k+\lambda+3)/(n-k+\lambda+2)^2,$$

the statement follows again from VII and (19). Therefore let us suppose

$$\alpha_{k-\lambda} > (n-k+\lambda+3)/(n-k+\lambda+2)^2. \tag{20}$$

By I and II we have

$$(n-k+1)\alpha_k + \lambda\alpha_{k-\lambda} < 1. \tag{21}$$

First we shall prove the relation

$$\alpha_k + \frac{n-k+\lambda+2}{n-k+\lambda+3}\,\alpha_{k-\lambda} < 2\alpha_{k-\lambda}. \tag{22}$$

As in the case where $\lambda = 1$ we obtain from (20) and (21)

$$\alpha_k + \frac{n-k+\lambda+2}{n-k+\lambda+3}\alpha_{k-\lambda} < \frac{1}{n-k+1} - \frac{\lambda}{n-k+1}\alpha_{k-\lambda} + \frac{n-k+\lambda+2}{n-k+\lambda+3}\alpha_{k-\lambda}$$

$$< \alpha_{k-\lambda}\,\frac{(n-k+\lambda+2)(2n-2k+\lambda+3) - \lambda(n-k+\lambda+3)}{(n-k+1)(n-k+\lambda+3)}$$

$$= \frac{2n^2 - 4nk + 2n\lambda + 7n + 2k^2 - 2k\lambda - 7k + 2\lambda + 6}{n^2 - 2nk + n\lambda + 4n + k^2 - k\lambda - 4k + \lambda + 3}\alpha_{k-\lambda} \leqq 2\alpha_{k-\lambda}$$

since $n \geqq k$.

Next, we state that under the assumption (20)

$$\text{(23)} \qquad \frac{n-k+\mu+2}{n-k+\mu+3}\alpha_{k-\mu} < \alpha_{k-\lambda} \qquad \text{for } \mu = 1, 2, \cdots, \lambda - 1.$$

As in (21) we have

$$(n-k+\mu+1)\alpha_{k-\mu} + (\lambda-\mu)\alpha_{k-\lambda} < 1.$$

Therefore, because of (20),

$$\text{(24)}\begin{cases} \alpha_{k-\mu} < \dfrac{1-(\lambda-\mu)\alpha_{k-\lambda}}{n-k+\mu+1} < \dfrac{(n-k+\lambda+2)^2 - (\lambda-\mu)(n-k+\lambda+3)}{(n-k+\mu+1)(n-k+\lambda+3)}\alpha_{k-\lambda} \\ \qquad = \dfrac{n^2-2nk+n\lambda+4n+k^2-k\lambda-4k+\lambda+4+\mu n-\mu k+\mu\lambda+3\mu}{n^2-2nk+n\lambda+4n+k^2-k\lambda-4k+\lambda+3+\mu n-\mu k+\mu\lambda+3\mu}\alpha_{k-\lambda}. \end{cases}$$

Putting

$$n - k + \mu + 3 = r$$

and

$$n^2 - 2nk + n\lambda + 4n + k^2 - k\lambda - 4k + \lambda + 3 + \mu n - \mu k + \mu\lambda + 3\mu = s$$

we obtain from (24)

$$\frac{n-k+\mu+2}{n-k+\mu+3}\alpha_{k-\mu}$$

$$= \frac{r-1}{r}\alpha_{k-\mu} < \frac{(r-1)(s+1)}{rs}\alpha_{k-\lambda} = \left(1 + \frac{r-s-1}{rs}\right)\alpha_{k-\lambda} < \alpha_{k-\lambda}$$

if

$$r < s + 1.$$

But this is true, since the following three inequalities hold

$$n - k \leqq n^2 - 2nk + k^2,$$

$$k(4 + \lambda + \mu) \leqq n(4 + \lambda + \mu),$$

$$\mu + 3 < 3\mu + 4 + \mu\lambda + \lambda.$$

Therefore (23) is proved.

Finally it follows from (22) and (23) that

$$\alpha_k + \frac{n-k+3}{n-k+4}\alpha_{k-1} + \cdots + \frac{n-k+\lambda+2}{n-k+\lambda+3}\alpha_{k-\lambda} < 2\alpha_{k-\lambda} + (\lambda - 1)\alpha_{k-\lambda}$$
$$= (\lambda+1)\alpha_{k-\lambda} \leqq \alpha_{k-\lambda,k}$$

because of II. Hence Theorem 2 is proved for $l = \lambda$ and consequently for each l.

4. Some lemmas

LEMMA 1: *Suppose $n = 2$ and*

$$\alpha_1 + \alpha_2 < 2/3. \tag{25}$$

We put

$$\gamma^* = \operatorname{Max}\left\{2\alpha_1,\ \alpha_2 + \frac{1}{1+[1/\alpha_1]},\ \alpha_1 + \frac{1}{1+[1/\alpha_2]}\right\}; \tag{26}$$

then we have

$$\alpha_1 + \alpha_2 - \gamma^* \leqq .05. \tag{27}$$

I.e., the difference between $\alpha_1 + \alpha_2$ and the lower bound for the density of $A_1 + A_2$ obtained by II and VI is less than .05, if (25) is satisfied.

PROOF: If (27) is not satisfied, we have

$$\operatorname{Min}\left\{\alpha_2 - \alpha_1,\ \alpha_1 - \frac{1}{1+[1/\alpha_1]},\ \alpha_2 - \frac{1}{1+[1/\alpha_2]}\right\} > .05 \tag{28}$$

because of (26). But the difference $1/k - 1/(k+1)$, where $k \geqq 4$ is an integer, is less than or equal to $\frac{1}{4} - \frac{1}{5} = \frac{1}{20}$. Hence

$$\alpha_1 - \frac{1}{1+[1/\alpha_1]} \leqq \frac{1}{[1/\alpha_1]} - \frac{1}{1+[1/\alpha_1]} \leqq \frac{1}{20} \qquad \text{for } \alpha_1 \leqq \tfrac{1}{4}. \tag{29}$$

Moreover (29) is true for $\frac{1}{4} < \alpha_1 \leqq \frac{1}{4} + \frac{1}{20} = .3$; therefore we have $\alpha_1 > .3$ and $\alpha_2 > .3 + .05 > \frac{1}{3}$ because of (28). Thus $\alpha_2 > \frac{1}{3} + .05$ again because of (28), and we obtain

$$\alpha_1 + \alpha_2 > .3 + \tfrac{1}{3} + .05 > \tfrac{2}{3}.$$

This contradicts (25). Therefore (27) is valid.

LEMMA 2: *Suppose $n = 4$ and $\gamma < 1$. Denote by γ^*_{123} the lower bound for the density of $A_1 + A_2 + A_3$ obtained by the theorems of this paper. Then we have*

$$d = \alpha_1 + \alpha_2 + \alpha_3 - \gamma^*_{123} \leqq \tfrac{7}{90}.$$

PROOF: We write

$$\delta_\nu = \alpha_\nu - \frac{1}{1+[1/\alpha_\nu]} \qquad \text{for } \nu = 1, 2, 3.$$

Then by VI

$$d \leqq \operatorname*{Min}_{\kappa \neq \lambda} (\delta_\kappa + \delta_\lambda). \tag{30}$$

It follows from Theorem 2 that the density α_{12} of $A_1 + A_2$ satisfies the inequality

$$\alpha_{12} \geqq \alpha_2 + \tfrac{5}{8}\alpha_1 . \tag{31}$$

On the other hand we have by I, II, and $\gamma < 1$

$$\alpha_{12} + 2\alpha_3 < 1, \tag{32}$$

hence by (31)

$$2\alpha_3 + \alpha_2 + \tfrac{5}{8}\alpha_1 < 1. \tag{33}$$

We have now to distinguish between some cases.

1) $$\alpha_1 \geqq \tfrac{121}{540}.$$

Because of Corollary 4 of Theorem 1 we have $\alpha_1 + \alpha_2 + \alpha_3 < \frac{3}{4}$ and by II

$$d \leqq \alpha_1 + \alpha_2 + \alpha_3 - 3\alpha_1 < \frac{135 - 121}{180} = \frac{7}{90}.$$

Therefore we may assume

$$\alpha_1 < \tfrac{121}{540} < \tfrac{7}{80},$$

$$\delta_1 \leqq \tfrac{1}{80}. \tag{34}$$

2) $$\alpha_2 \leqq \tfrac{1}{5} + \tfrac{2}{45} = \tfrac{11}{45}.$$

We have by (30) and (34)

$$d \leqq \delta_2 + \delta_1 \leqq \tfrac{2}{45} + \tfrac{1}{80} = \tfrac{7}{90}.$$

Hence we may assume

$$\alpha_2 > \tfrac{11}{45} \tag{35}$$

and because of (33)

$$\alpha_3 < (1 - \alpha_2)/2 = \tfrac{17}{45}. \tag{36}$$

3) $$\alpha_3 > \tfrac{1}{3}.$$

It follows from (33) and (35) that

$$\alpha_1 < \tfrac{6}{5}(1 - 2\alpha_3 - \alpha_2) < \tfrac{6}{5}(1 - \tfrac{2}{3} - \tfrac{11}{45}) = \tfrac{8}{75} < \tfrac{1}{9}.$$

Therefore we have $\delta_1 < \frac{1}{90}$ and $\delta_3 < \frac{17}{45} - \frac{1}{3} = \frac{2}{45}$ because of (36), hence $d < \frac{7}{90}$ because of (30). Therefore we may assume

$$\alpha_3 \leqq \tfrac{1}{3}. \tag{37}$$

4) $$\alpha_2 \geqq \tfrac{13}{45} \qquad \text{or} \qquad \alpha_3 \leqq \tfrac{13}{45}.$$

We have by II, (37), (35), and (34)

$$d \leqq \alpha_3 - \alpha_2 + \delta_1 \leqq \tfrac{2}{45} + \tfrac{1}{30} = \tfrac{7}{90}.$$

Hence we may assume

$$\alpha_2 < \tfrac{13}{45}, \tag{38}$$

$$\alpha_3 > \tfrac{13}{45} > \tfrac{1}{4}. \tag{39}$$

5) $$\alpha_3 \leqq \tfrac{1}{4} + \tfrac{2}{45}.$$

We have by (34) and (39)

$$d \leqq \delta_1 + \delta_3 \leqq \tfrac{1}{30} + \tfrac{2}{45} = \tfrac{7}{90}.$$

Hence we may assume

$$\alpha_3 > \tfrac{53}{180}. \tag{40}$$

6) $$\alpha_1 \leqq \tfrac{7}{36}.$$

We have $\delta_1 \leqq \frac{1}{36}$ and $\delta_2 \leqq \frac{1}{20}$ because of (38). Therefore we may assume

$$\alpha_1 > \tfrac{7}{36}$$

and obtain by VI

$$\alpha_{12} \geqq \alpha_2 + \tfrac{1}{6},$$

$$\alpha_{12} + 2\alpha_3 > \tfrac{1}{6} + \tfrac{11}{45} + \tfrac{53}{90} = 1$$

because of (35) and (40). This contradicts (32) and Lemma 2 is proved.

5. Estimates for the sum of the densities in the cases $n = 3$ and $n = 4$

THEOREM 3: *Let $n = 3$. If $A_1 + A_2 + A_3$ is not the set of all positive integers, then we have*

$$\alpha_1 + \alpha_2 + \alpha_3 < 1.05. \tag{41}$$

PROOF: Let γ_{12}^* be the lower bound for the density of $A_1 + A_2$ obtained by II and VI. Because of Theorem 1 we have $\alpha_1 + \alpha_2 < \frac{2}{3}$. Therefore it follows from Lemma 1 that

$$\alpha_1 + \alpha_2 - \gamma_{12}^* < .05. \tag{42}$$

On the other hand we have by I

$$\gamma_{12}^* + \alpha_3 < 1. \tag{43}$$

Adding (42) and (43) we obtain (41).

THEOREM 4: *Let $n = 4$ be. If $A_1 + A_2 + A_3 + A_4$ is not the set of all positive integers, then we have*

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 < \tfrac{97}{90} < 1.0778. \tag{44}$$

PROOF: Let γ_{123}^* be the lower bound for the density of $A_1 + A_2 + A_3$ obtained by the theorems of this paper. It follows from Lemma 2 that

$$\alpha_1 + \alpha_2 + \alpha_3 - \gamma_{123}^* < \tfrac{7}{90} \tag{45}$$

and from I that

$$\gamma_{123}^* + \alpha_4 < 1. \tag{46}$$

By the addition of (45) and (46) we obtain (44).

6. Improvement of the Theorems V, X, and XI

THEOREM 5: *If the densities* $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$ *of* $A_1, A_2, \cdots, A_n$ *satisfy the condition*

$$\alpha_n \geqq 1 - n\alpha_1, \tag{47}$$

then we have

$$\gamma \geqq \frac{n-1}{n} + \frac{\alpha_n}{n};$$

if

$$\alpha_1 \geqq 1 - n\alpha_2, \tag{48}$$

then

$$\gamma \geqq \frac{n-1}{n} + \frac{\alpha_1}{n}.$$

PROOF: If (47) is satisfied, we set

$$\alpha = (1 - \alpha_n)/n, \tag{49}$$

then $\alpha < 1/n$ and by (47)

$$\alpha_\nu \geqq \alpha_1 \geqq (1 - \alpha_n)/n = \alpha \quad (\nu = 2, 3, \cdots, n-1),$$
$$\alpha_n = 1 - n\alpha.$$

It follows that for $x = 1, 2, \cdots$

$$A_\nu(x) \geqq \alpha_\nu x \geqq \alpha x \geqq \alpha x - (\nu - 1)/n \qquad (\nu = 1, 2, \cdots, n-1),$$
$$A_n(x) \geqq (1 - n\alpha)x.$$

Applying III for $\mu = 0$ to the sets $A_1, A_2, \cdots, A_n$ we obtain by (49)

$$\gamma \geqq 1 - \alpha = 1 - \frac{1 - \alpha_n}{n} = \frac{n-1}{n} + \frac{\alpha_n}{n}.$$

Suppose now that (48) is satisfied. Here we set

$$\alpha = (1 - \alpha_1)/n. \tag{50}$$

Then we have $\alpha < 1/n$ and

$$\alpha_\nu \geqq \alpha_2 \geqq (1 - \alpha_1)/n = \alpha \qquad (\nu = 3, 4, \cdots, n),$$

$$\alpha_1 = 1 - n\alpha.$$

Applying III to the sets $A_2, A_3, \cdots, A_n, A_1$ we obtain by (50)

$$\gamma \geqq 1 - \alpha = 1 - \frac{1 - \alpha_1}{n} = \frac{n-1}{n} + \frac{\alpha_1}{n}.$$

By means of Theorem 5 it is possible to improve XI. First we consider the case $n = 2$.

Theorem 6: *For $n = 2$ we have $\gamma \geqq \frac{9}{10}(\alpha_1 + \alpha_2)$.*

Proof: We have to distinguish between certain cases:

1) $$\alpha_1 \leqq \tfrac{5}{16}.$$

It follows from VII that

$$\gamma \geqq \alpha_2 + \tfrac{4}{5}\alpha_1$$

and from II that

$$\gamma \geqq 2\alpha_1,$$

hence

$$10\gamma + \gamma \geqq 10(\alpha_1 + \alpha_2),$$

$$\gamma \geqq \tfrac{10}{11}(\alpha_1 + \alpha_2) > \tfrac{9}{10}(\alpha_1 + \alpha_2).$$

Therefore we may assume that

(51) $$\tfrac{1}{2} > \alpha_1 > \tfrac{5}{16}.$$

2) $$\alpha_1 \geqq \tfrac{9}{11}\alpha_2.$$

We deduce from II

$$\gamma \geqq 2\alpha_1 = \tfrac{9}{10}\alpha_1 + \tfrac{11}{10}\alpha_1 \geqq \tfrac{9}{10}(\alpha_1 + \alpha_2).$$

Hence we may assume that

(52) $$\alpha_1 < \tfrac{9}{11}\alpha_2$$

and, because of (51)

(53) $$\alpha_2 > \tfrac{11}{9} \cdot \tfrac{5}{16} = \tfrac{55}{144}.$$

3) $$9\alpha_1 + 4\alpha_2 \leqq 5.$$

It follows from (51) and (53) that

$$2\alpha_1 + \alpha_2 \geqq \tfrac{145}{144} > 1,$$

hence we have because of Theorem 5 for $n = 2$

$$\gamma \geqq \tfrac{1}{2} + \frac{\alpha_2}{2} \geqq \tfrac{9}{10}\alpha_1 + \tfrac{4}{10}\alpha_2 + \tfrac{1}{2}\alpha_2 = \tfrac{9}{10}(\alpha_1 + \alpha_2).$$

Therefore we may assume

(54) $$9\alpha_1 + 4\alpha_2 > 5.$$

4) $$\alpha_1 \leqq \tfrac{1}{3}.$$

By VI and (51) we have

(55) $$\gamma \geqq \tfrac{1}{4} + \alpha_2 = \tfrac{1}{4} + \tfrac{1}{10}\alpha_2 + \tfrac{9}{10}\alpha_2 .$$

On the other hand, it follows from (54) that

$$\tfrac{1}{10}\alpha_2 > \tfrac{1}{8} - \tfrac{9}{40}\alpha_1 ,$$

hence by (55)

$$\gamma \geqq \tfrac{3}{8} - \tfrac{9}{40}\alpha_1 + \tfrac{9}{10}\alpha_2 \geqq \tfrac{9}{8}\alpha_1 - \tfrac{9}{40}\alpha_1 + \tfrac{9}{10}\alpha_2 = \tfrac{9}{10}(\alpha_1 + \alpha_2).$$

5) $$\tfrac{1}{3} < \alpha_1 < \tfrac{3}{7}.$$

By VI and (52) we have

$$\gamma \geqq \tfrac{1}{3} + \alpha_2 = \tfrac{1}{3} + \tfrac{1}{10}\alpha_2 + \tfrac{9}{10}\alpha_2 \geqq \tfrac{1}{3} + \tfrac{11}{90}\alpha_1 + \tfrac{9}{10}\alpha_2$$
$$> \tfrac{7}{9}\alpha_1 + \tfrac{11}{90}\alpha_1 + \tfrac{9}{10}\alpha_2 = \tfrac{9}{10}(\alpha_1 + \alpha_2).$$

6) $$\alpha_1 \geqq \tfrac{3}{7}.$$

It follows from (52) that $\alpha_2 > \tfrac{11}{21}$. Because of IV we have

$$\gamma \geqq \frac{\alpha_2}{1 - \alpha_1} > \frac{11 \cdot 7}{21 \cdot 4} > \tfrac{9}{10} > \tfrac{9}{10}(\alpha_1 + \alpha_2).$$

This proves Theorem 6.

THEOREM 7: $\gamma \geqq \alpha_1 + \tfrac{5}{8}\alpha_2$.

PROOF: It follows from Theorem 6 that

$$\gamma \geqq \tfrac{9}{10}\alpha_1 + \tfrac{9}{10}\alpha_2 \geqq \alpha_1 + \tfrac{5}{8}\alpha_2$$

if $\alpha_2 \geqq \tfrac{3}{8}\alpha_1$. Therefore we may assume

(56) $$\alpha_2 < \tfrac{3}{8}\alpha_1.$$

2) It follows from II that

$$\gamma \geqq 2\alpha_1 \geqq \alpha_1 + \tfrac{5}{8}\alpha_2$$

if $\alpha_2 \leqq \tfrac{8}{5}\alpha_1$; hence we may assume

(57) $$\alpha_2 > \tfrac{8}{5}\alpha_1.$$

3) Theorem VI implies

$$\gamma \geqq \alpha_2 + \frac{\alpha_1}{1 + \alpha_1} \geqq \alpha_1 + \tfrac{5}{6}\alpha_2$$

if $\alpha_2 \geqq 6\alpha_1^2/(1 + \alpha_1)$. Therefore we may assume

$$6\alpha_1^2/(1 + \alpha_1) > \alpha_2 > \tfrac{6}{5}\alpha_1$$

because of (57), hence $\alpha_1 > \frac{1}{4}$.

4) Since $\alpha_1 > \frac{1}{4}$, we have

$$\gamma \geqq \alpha_2 + \tfrac{1}{4} \geqq \alpha_1 + \tfrac{5}{6}\alpha_2$$

if $\alpha_2 \geqq 6\alpha_1 - \frac{3}{2}$. Therefore we may assume

$$6\alpha_1 - \tfrac{3}{2} > \alpha_2 > \tfrac{6}{5}\alpha_1$$

because of (57), hence $\alpha_1 > \frac{5}{16}$, $\alpha_2 > \frac{3}{8}$.

5) Now we have $2\alpha_1 + \alpha_2 > 1$, hence by Theorem 5

$$\gamma \geqq \tfrac{1}{2} + \tfrac{1}{2}\alpha_2 \geqq \alpha_1 + \tfrac{5}{6}\alpha_2$$

if $\alpha_2 \leqq \frac{3}{2} - 3\alpha_1$. Therefore we may assume

$$\tfrac{3}{2} - 3\alpha_1 < \alpha_2 < \tfrac{3}{2}\alpha_1$$

because of (56), hence $\alpha_1 > \frac{1}{3}$.

6) Since $\alpha_1 > \frac{1}{3}$, we have

$$\gamma \geqq \alpha_2 + \tfrac{1}{3} \geqq \alpha_1 + \tfrac{5}{6}\alpha_2$$

if $\alpha_1 \leqq \frac{1}{6}\alpha_2 + \frac{1}{3}$. Therefore we may assume

$$\tfrac{1}{6}\alpha_2 + \tfrac{1}{3} < \alpha_1 < \tfrac{5}{6}\alpha_2$$

because of (57), hence $\alpha_2 > \frac{1}{2}$.

7) Since $\alpha_2 > \frac{1}{2}$, it follows for $\alpha_2 \leqq \frac{3}{5}$ that

$$\gamma \geqq \alpha_1 + \tfrac{1}{2} \geqq \alpha_1 + \tfrac{5}{6}\alpha_2 .$$

Hence we may assume $\alpha_2 > \frac{3}{5}$ and $\alpha_1 > \frac{3}{5}$ because of (56). This gives a contradiction to I, and our theorem is proved.

Before improving XI for $n > 2$ we have to prove the following

Lemma 3: *Suppose* $n = 3$ *and*

$$\alpha_3 \geqq \alpha_2 \geqq \alpha_1 > .24 \tag{58}$$

If γ_{123}^*, d, *and* δ_ν *have the same significance as in Lemma* 2, *then we have* $d \leqq \frac{22}{225}$.

Proof: Let $\gamma_{\kappa\lambda}^*$ be the lower bound for the density of $A_\kappa + A_\lambda$ ($\kappa \neq \lambda$) obtained by the theorems of this paper. We set

$$d_{\kappa\lambda} = \alpha_\kappa + \alpha_\lambda - \gamma_{\kappa\lambda}^* .$$

Since $\alpha_1 + 2\alpha_2 < 1$, we obtain from (58)

$$(59) \qquad \alpha_2 < .38.$$

We have again to distinguish between several cases:

1) $$\tfrac{1}{3} < \alpha_2 < .38.$$

Suppose first $\alpha_1 > \frac{1}{4}$. Then we have by Theorem 1 for $m = 2$

$$d \leqq \delta_1 + \delta_2 \leqq \alpha_1 - \tfrac{1}{4} + \alpha_2 - \tfrac{1}{3} = \alpha_1 + \alpha_2 - \tfrac{7}{12} \leqq \tfrac{2}{3} - \tfrac{7}{12} < \tfrac{22}{225}.$$

Suppose next $\alpha_1 \leqq \frac{1}{4}$; then by (59)

$$d \leqq \delta_1 + \delta_2 < \tfrac{1}{20} + \tfrac{7}{150} = \tfrac{29}{300} < \tfrac{22}{225}.$$

Because of (58) and (59) we may assume

$$(60) \qquad .24 < \alpha_2 \leqq \tfrac{1}{3}.$$

2) $$\alpha_3 \geqq \tfrac{857}{1800}.$$

Because of (58) we have $\gamma_{12}^* \geqq 2\alpha_1 > .48$ and

$$\gamma_{123}^* \geqq 2 \min (\gamma_{12}^*, \alpha_3) \geqq 2 \min (\tfrac{48}{100}, \tfrac{857}{1800}) = \tfrac{857}{900}.$$

On the other hand we have by Theorem 3

$$\alpha_1 + \alpha_2 + \alpha_3 < 1.05,$$

hence

$$d < 1.05 - \tfrac{857}{900} = \tfrac{22}{225}.$$

Therefore we may assume

$$(61) \qquad \alpha_3 < \tfrac{857}{1800}.$$

3) $$\alpha_2 \leqq \tfrac{1}{4}.$$

a) $$\alpha_3 \leqq \tfrac{1}{3} + \tfrac{79}{900}.$$

Then we have, because of (58),

$$\delta_3 \leqq \max (\tfrac{1}{12}, \tfrac{79}{900}) = \tfrac{79}{900}; \qquad d_{12} \leqq \alpha_2 - \alpha_1 \leqq \tfrac{1}{100}; \qquad d \leqq d_{12} + \delta_3 \leqq \tfrac{22}{225}.$$

b) $$\alpha_3 > \tfrac{1}{3} + \tfrac{79}{900}.$$

Because of (58) and (61) we have $\gamma_{12}^* \geqq 2\alpha_1 > \alpha_3$, hence $\gamma_{123}^* \geqq 2\alpha_3$, and

$$d \leqq \alpha_1 + \alpha_2 + \alpha_3 - 2\alpha_3 \leqq 2\alpha_2 - \alpha_3 < \tfrac{1}{2} - \tfrac{379}{900} = \tfrac{71}{900} < \tfrac{22}{225}.$$

Therefore the case 3) is also proved and we may assume that $\alpha_2 > \frac{1}{4}$.

4) $$\tfrac{1}{4} < \alpha_2 \leqq \tfrac{1}{4} + \tfrac{43}{900} = \tfrac{67}{225}.$$

We have $\delta_2 \leqq \frac{43}{900}$; $\delta_1 \leqq \frac{1}{20}$; $d \leqq \frac{22}{225}$. Hence we may assume $\alpha_2 > \frac{67}{225}$.

5) $$\tfrac{67}{225} < \alpha_2 \leqq .3.$$

a) $$\alpha_3 \leqq \tfrac{343}{900}.$$

If $\alpha_3 \leqq \frac{1}{3}$, we have $d_{23} \leqq \alpha_3 - \alpha_2 < \frac{32}{900}$. If $\alpha_3 > \frac{1}{3}$, we have $d_{23} \leqq \delta_3 = \frac{343}{900} - \frac{1}{3} = \frac{43}{900}$. Hence

$$d \leqq \delta_1 + d_{23} \leqq \tfrac{1}{20} + \tfrac{43}{900} = \tfrac{22}{225},$$

since $\alpha_1 \leqq \alpha_2 \leqq \frac{3}{10}$. Therefore we assume

(62) $$\alpha_3 > \tfrac{343}{900}.$$

b) $$\alpha_1 \leqq \tfrac{1}{4} \quad \text{and} \quad \alpha_3 \leqq \tfrac{1}{3} + \tfrac{7}{75}.$$

From (58) and (62) it follows that $\alpha_1 + 2\alpha_3 > 1$. Hence we obtain by Theorem 5

$$\overset{*}{\gamma}_{13} \geqq \tfrac{1}{2} + \tfrac{1}{2}\alpha_1$$

and by IV and (58)

$$\overset{*}{\gamma}_{123} \geqq \frac{\overset{*}{\gamma}_{13}}{1 - \alpha_2} \geqq \frac{1 + \alpha_1}{2(1 - \alpha_2)} \geqq \frac{1.24 \cdot 225}{2 \cdot 158} = \frac{279}{316},$$

$$d \leqq \alpha_1 + \alpha_2 + \alpha_3 - \tfrac{279}{316} < \tfrac{1}{4} + \tfrac{3}{10} + \tfrac{1}{3} + \tfrac{7}{75} - \tfrac{279}{316} = \tfrac{293}{300} - \tfrac{279}{316} < \tfrac{22}{225}.$$

c) $$\alpha_1 \leqq \tfrac{1}{4} \quad \text{and} \quad \alpha_3 > \tfrac{1}{3} + \tfrac{7}{75}.$$

We have

$$2\alpha_2 + \alpha_3 > \tfrac{134}{225} + \tfrac{1}{3} + \tfrac{7}{75} > 1,$$

hence by Theorem 5 and by (58)

$$\overset{*}{\gamma}_{123} \geqq \tfrac{1}{2} + \tfrac{1}{2}\alpha_3 + \tfrac{1}{5},$$

$$d = \alpha_1 + \alpha_2 + \alpha_3 - \overset{*}{\gamma}_{123} \leqq \tfrac{1}{4} + \tfrac{3}{10} + \tfrac{857}{3600} - \tfrac{1}{2} - \tfrac{1}{5} = \tfrac{317}{3600} < \tfrac{22}{225}$$

because of (61).

d) $$\tfrac{1}{4} < \alpha_1 \leqq \tfrac{1}{4} + \tfrac{43}{900}.$$

We have

$$\delta_1 \leqq \tfrac{43}{900}, \qquad \delta_2 \leqq \tfrac{1}{20}, \qquad d \leqq \tfrac{22}{225}.$$

e) $$\alpha_1 > \tfrac{1}{4} + \tfrac{43}{900} = \tfrac{67}{225} \quad \text{and} \quad \alpha_3 \leqq \tfrac{128}{395} + \tfrac{22}{225}.$$

It follows similarly as above in 5*b* that

$$\overset{*}{\gamma}_{123} \geqq \frac{1 + \alpha_1}{2(1 - \alpha_2)} \geqq \frac{73}{79},$$

$$d \leqq \alpha_1 + \alpha_2 + \alpha_3 - \tfrac{73}{79} < \tfrac{3}{10} + \tfrac{3}{10} + \tfrac{128}{395} + \tfrac{22}{225} - \tfrac{73}{79} = \tfrac{22}{225}.$$

f) $$\alpha_1 > \tfrac{67}{225} \quad \text{and} \quad \alpha_3 > \tfrac{128}{395} + \tfrac{22}{225}.$$

We have

$$2\alpha_2 + \alpha_3 > \tfrac{134}{225} + \tfrac{128}{395} + \tfrac{22}{225} = \tfrac{52}{75} + \tfrac{128}{395} > \tfrac{4}{5}(\tfrac{13}{15} + \tfrac{2}{5}) > 1,$$

hence by Theorem 5

$$\gamma_{123}^{*} \geqq \tfrac{1}{2} + \tfrac{1}{2}\alpha_3 + \tfrac{1}{4}$$

and by (61)

$$d < \tfrac{3}{5} + \tfrac{857}{3600} - \tfrac{3}{4} = \tfrac{317}{3600} < \tfrac{22}{225}.$$

Therefore the lemma is proved also in the case 5) and we assume

$$\tfrac{3}{10} < \alpha_2 \leqq \tfrac{1}{3}. \tag{63}$$

6) $$\alpha_1 + \alpha_2 \leqq \tfrac{1}{2} + \tfrac{22}{225} = \tfrac{269}{450}.$$

It follows from (63) and (58) that

$$\gamma_{12}^{*} \geqq \alpha_2 + \tfrac{1}{5} > \tfrac{1}{2};$$

hence, because of VI,

$$\gamma_{123}^{*} \geqq \alpha_3 + \tfrac{1}{2},$$

$$d = \alpha_1 + \alpha_2 + \alpha_3 - \gamma_{123}^{*} \leqq \tfrac{1}{2} + \tfrac{22}{225} + \alpha_3 - \alpha_3 - \tfrac{1}{2} = \tfrac{22}{225}.$$

Therefore we may assume

$$\alpha_1 + \alpha_2 > \tfrac{269}{450} \tag{64}$$

and, because of (63),

$$\alpha_1 > \tfrac{269}{450} - \tfrac{1}{3} = \tfrac{119}{450} > \tfrac{1}{4}. \tag{65}$$

7) $$\alpha_3 \leqq \tfrac{1}{3}.$$

a) $$\alpha_1 \geqq \tfrac{64}{225}.$$

By II we have

$$d \leqq \alpha_1 + \alpha_2 + \alpha_3 - 3\alpha_1 \leqq 2\alpha_3 - 2\alpha_1 \leqq \tfrac{2}{3} - \tfrac{128}{225} = \tfrac{22}{225}.$$

b) $$\alpha_1 < \tfrac{64}{225}.$$

Because of (65) we have

$$\delta_1 < \tfrac{31}{900}, \qquad d_{23} \leqq \alpha_3 - \alpha_2 \leqq \tfrac{1}{3} - \tfrac{3}{10} = \tfrac{1}{30}, \qquad d < \tfrac{61}{900} < \tfrac{22}{225}.$$

Therefore we may assume $\alpha_3 > \frac{1}{3}$.

8) $$\tfrac{1}{3} < \alpha_3 \leqq \tfrac{1}{3} + \tfrac{101}{1800} = \tfrac{701}{1800}.$$

For $\alpha_1 \leqq \frac{1}{4} + \frac{1}{24}$ we have by (65)

$$d_{12} \leqq \delta_1 \leqq \tfrac{1}{24}. \tag{66}$$

For $\alpha_1 > \frac{1}{4} + \frac{1}{24}$ the inequality (66) is also valid, since we have

$$d_{12} \leqq \alpha_2 - \alpha_1 \leqq \tfrac{1}{3} - \tfrac{7}{24} = \tfrac{1}{24}.$$

Hence we obtain

$$d \leqq d_{12} + \delta_3 \leqq \tfrac{1}{24} + \tfrac{101}{1800} = \tfrac{22}{225}$$

and we may assume

$$\alpha_3 > \tfrac{701}{1800}. \tag{67}$$

9) It follows from (66) and IV that

$$\gamma_{12}^{*} \geqq \alpha_1 + \alpha_2 - \tfrac{1}{24},$$

$$\gamma_{123}^{*} \geqq (\alpha_1 + \alpha_2 - \tfrac{1}{24})/(1 - \alpha_3),$$

$$\begin{aligned} d &\leqq \alpha_1 + \alpha_2 + \alpha_3 - \frac{\alpha_1 + \alpha_2 - \frac{1}{24}}{1 - \alpha_3} \\ &= \alpha_3 - \frac{\alpha_3(\alpha_1 + \alpha_2) - \frac{1}{24}}{1 - \alpha_3} \leqq \alpha_3 - (\tfrac{269}{450}\alpha_3 - \tfrac{1}{24})(1 - \alpha_3)^{-1} \end{aligned} \tag{68}$$

because of (64). If we write

$$f(\alpha_3) = \alpha_3 - (\tfrac{269}{450}\,\alpha_3 - \tfrac{1}{24})(1 - \alpha_3)^{-1},$$

we have

$$f'(\alpha_3) = 1 - \{\tfrac{269}{450}\,(1 - \alpha_3) + \tfrac{269}{450}\,\alpha_3 - \tfrac{1}{24}\}(1 - \alpha_3)^{-2} = 1 - \tfrac{1001}{1800}\,(1 - \alpha_3)^{-2}.$$

Therefore $f(\alpha_3)$ has its maximum in the interval $0 \leqq \alpha_3 < 1$ for

$$\alpha_3 = 1 - (\tfrac{1001}{1800})^{\frac{1}{2}} < 1 - \tfrac{31}{43} < \tfrac{701}{1800}.$$

Hence $f(\alpha_3)$ is decreasing in the interval $\frac{701}{1800} < \alpha_3 < 1$ and it follows from (67) and (68) that

$$\begin{aligned} d &\leqq \frac{701}{1800} - \left(\frac{701}{1800}\cdot\frac{269}{450} - \frac{1}{24}\right)\cdot\frac{1800}{1099} \\ &= \frac{701\cdot 1099 - 701\cdot 1076 + 75\cdot 1800}{1800\cdot 1099} < \frac{701 + 5400}{1800\cdot 40} < \frac{22}{225}. \end{aligned}$$

This proves Lemma 3.

Using this lemma we can improve XI.

Theorem 8: *Write* $\alpha_1 + \alpha_2 + \cdots + \alpha_n = \sigma_n$; *then we have*

$$\gamma \geqq \tfrac{9}{10}\,\sigma_2; \qquad \gamma \geqq \tfrac{81}{92}\,\sigma_3 > .8804\sigma_3; \qquad \gamma \geqq \tfrac{972}{1117}\,\sigma_4 > .8701\sigma_4;$$

$$\gamma \geqq \tfrac{84029}{96817}\,\sigma_n > .8679\sigma_n \qquad \textit{for } n \geqq 5.$$

Proof: The case $n = 2$ is already proved by Theorem 6.

1) $$n = 3$$

a) $$\alpha_1 \leqq \tfrac{6}{25}.$$

Denote by α_{23} the density of $A_2 + A_3$. It follows from Theorem 6 that

$$\alpha_{23} \geqq \tfrac{9}{10}(\alpha_2 + \alpha_3),$$

hence from Theorem VII for $k = 6$ and from II

(69) $$\gamma \geqq \tfrac{9}{10}(\alpha_2 + \alpha_3) + \tfrac{5}{6}\alpha_1,$$

(70) $$\gamma \geqq 3\alpha_1,$$

and, because of (69) and (70),

$$90\gamma + 2\gamma \geqq 81(\alpha_2 + \alpha_3) + 75\alpha_1 + 6\alpha_1,$$

$$\gamma \geqq \tfrac{81}{92}\sigma_3.$$

b) $$\alpha_1 > \tfrac{6}{25}.$$

Suppose

(71) $$\gamma < \tfrac{81}{92}\sigma_3.$$

Then we have

$$\tfrac{18}{25} < 3\alpha_1 \leqq \gamma < \tfrac{81}{92}\sigma_3,$$

(72) $$\sigma_3 > \tfrac{184}{225}.$$

On the other hand it follows from Lemma 3 and (71) that

$$\sigma_3 - \tfrac{22}{225} \leqq \overset{*}{\gamma}_{123} \leqq \gamma < \tfrac{81}{92}\sigma_3,$$

$$\tfrac{11}{92}\sigma_3 < \tfrac{22}{225},$$

$$\sigma_3 < \tfrac{184}{225}.$$

This is incompatible with (72). Therefore the supposition (71) is impossible and in both cases we have

(73) $$\gamma \geqq \tfrac{81}{92}\sigma_3.$$

2) $$n = 4.$$

It is possible to treat this case as we have treated the case $n = 3$. But the proof of the lemma analogous to the Lemma 3 is troublesome. Therefore we use a method which is simpler, but less sharp. Denote by α_{234} the density of $A_2 + A_3 + A_4$. It follows from (73) that

$$\alpha_{234} \geqq \tfrac{81}{92}(\alpha_2 + \alpha_3 + \alpha_4),$$

hence from VIII and from II

(74) $$\gamma \geqq \tfrac{81}{92}(\alpha_2 + \alpha_3 + \alpha_4) + \tfrac{5}{6}\alpha_1,$$

(75) $$\gamma \geqq 4\alpha_1.$$

Therefore we have by (74) and (75)

$$1104\gamma + 13\gamma \geqq 972\,(\alpha_2 + \alpha_3 + \alpha_4) + 920\alpha_1 + 52\alpha_1,$$

$$\gamma \geqq \tfrac{972}{1117}\,\sigma_4.$$

3) $$n = 5.$$

In similar way we obtain

$$39095\gamma + 102\gamma \geqq 39095\,\{\tfrac{972}{1117}(\alpha_2 + \alpha_3 + \alpha_4 + \alpha_5) + \tfrac{6}{7}\,\alpha_1\} + 510\alpha_1,$$

$$\gamma \geqq \tfrac{34020}{39197}\,\sigma_5.$$

4) $$n > 5.$$

It follows from VIII that

$$\gamma \geqq \tfrac{34020}{39197}\,(\alpha_n + \alpha_{n-1} + \alpha_{n-2} + \alpha_{n-3} + \alpha_{n-4}) + \tfrac{7}{8}\,\alpha_{n-5} + \cdots + \frac{n+1}{n+2}\,\alpha_1,$$

hence, because of $\tfrac{34020}{39197} < \tfrac{7}{8}$,

$$\gamma \geqq \tfrac{34020}{39197}\,\sigma_n.$$

7. Estimates for the sum of the densities for $n \geqq 5$

In order to obtain an estimate for the sum of the densities for $n = 5$ we could apply a similar method as in 5). But the proof of the lemma analogous to the Lemmas 1 and 2 would be very complicate. Therefore we are content to prove the following

THEOREM 9: *Suppose that $A_1 + A_2 + \cdots + A_n$ is not the set of all positive integers. Then we have*

$$\sigma_5 < \tfrac{43}{39} < 1.1026 \text{ and } \sigma_n < \frac{39197}{34020n + 5177} < 1.1522 \qquad \text{for } n > 5.$$

PROOF: For $n = 5$ we obtain from Theorem 2 for $k = 2$ and $l = 1$

(76) $$\alpha_{12} \geqq \alpha_2 + \tfrac{6}{7}\,\alpha_1$$

and from II and I

(77) $$\alpha_{345} \geqq 3\alpha_3,$$

(78) $$\alpha_{12} + \alpha_{345} < 1,$$

hence, because of (76), (77), and (78)

(79) $$\tfrac{6}{7}\,\alpha_1 + \alpha_2 + 3\alpha_3 < 1.$$

Moreover we have

(80) $$\alpha_1 + 4\alpha_2 < 1,$$

(81) $$2\alpha_1 + 3\alpha_3 < 1,$$

and finally by Theorem 2

$$\alpha_{1234} \geqq \alpha_4 + \tfrac{4}{5}\alpha_3 + \tfrac{5}{6}\alpha_2 + \tfrac{6}{7}\alpha_1,$$

hence, because of I

$$\alpha_5 + \alpha_4 + \tfrac{4}{5}\alpha_3 + \tfrac{5}{6}\alpha_2 + \tfrac{6}{7}\alpha_1 \leqq \alpha_5 + \alpha_{1234} < 1. \tag{82}$$

If we multiply (79), (80), (81), (82) by $\frac{3}{130}$, $\frac{14}{390}$, $\frac{17}{390}$, 1 respectively and add these inequalities, we obtain

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 < \tfrac{3}{130} + \tfrac{14}{390} + \tfrac{17}{390} + 1 = \tfrac{43}{39} < 1.1026.$$

For $n > 5$ we are content to prove an inequality less sharp. It follows from Theorem 8 and from II that

$$\tfrac{34020}{39197}(\alpha_1 + \alpha_2 + \cdots + \alpha_{n-1}) + \alpha_n < 1,$$

and because of $\alpha_n \geqq (\alpha_1 + \alpha_2 + \cdots + \alpha_n)/n$

$$\frac{34020}{39197}(\alpha_1 + \alpha_2 + \cdots + \alpha_{n-1} + \alpha_n) + \frac{5177}{39197}\,\frac{\alpha_1 + \alpha_2 + \cdots + \alpha_n}{n} < 1,$$

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n < \frac{39197n}{34020n + 5177} < 1.1522.$$

8. Improvement of a theorem of Schur

THEOREM 10: *If we set $c_n^{(1)} = n$ and for $m > 1$*

$$c_n^{(m)} = \frac{n(n+2)c_{n-1}^{(m-1)}}{(n+2)c_{n-1}^{(m-1)} + n^2 + n - 1} \qquad \text{if } c_{n-1}^{(m-1)} > \frac{n+1}{n+2}, \tag{83}$$

$$c_n^{(m)} = c_{n-1}^{(m-1)} \qquad \text{if } c_{n-1}^{(m-1)} \leqq \frac{n+1}{n+2}, \tag{84}$$

then we have

$$\gamma \geqq c_n^{(m)}(\alpha_1 + \alpha_2 + \cdots + \alpha_m).$$

PROOF: For $m = 1$ the assertion follows from II. Therefore we may assume that the theorem is proved for $m - 1$ and every n. Denoting the density of $A_2 + A_3 + \cdots + A_n$ by β we have

$$\beta \geqq c_{n-1}^{(m-1)}(\alpha_2 + \alpha_3 + \cdots + \alpha_m). \tag{85}$$

We have now to distinguish between two cases:

a) $$c_{n-1}^{(m-1)} \leqq \frac{n+1}{n+2}.$$

It follows from VIII and from (85) that

$$\gamma \geqq \beta + \frac{n+1}{n+2}\alpha_1 \geqq c_{n-1}^{(m-1)}(\alpha_2 + \alpha_3 + \cdots + \alpha_m) + \frac{n+1}{n+2}\alpha_1$$
$$\geqq c_{n-1}^{(m-1)}(\alpha_1 + \alpha_2 + \cdots + \alpha_m),$$

hence, because of (84),

$$\gamma \geqq c_n^{(m)}(\alpha_1 + \alpha_2 + \cdots + \alpha_m).$$

b)
$$c_{n-1}^{(m-1)} > \frac{n+1}{n+2}.$$

If we set

$$a = \frac{n(n+2)}{c_{n-1}^{(m-1)}(n+2) + n^2 + n - 1},$$

we have

$$a < \frac{n(n+2)}{n+1+n^2+n-1} = 1.$$

Therefore applying IX we obtain

$$\gamma \geqq a\beta + \alpha_1\left\{n - \frac{n^2+n-1}{n+2}\,a\right\}$$
$$= \frac{n(n+2)\beta}{c_{n-1}^{(m-1)}(n+2) + n^2 + n - 1} + \alpha_1\left\{n - \frac{n(n^2+n-1)}{c_{n-1}^{(m-1)}(n+2) + n^2 + n - 1}\right\}$$
$$= \frac{n}{c_{n-1}^{(m-1)}(n+2) + n^2 + n - 1}\left\{(n+2)\beta + \alpha_1 c_{n-1}^{(m-1)}(n+2)\right\}$$
$$\geqq \frac{n(n+2)c_{n-1}^{(m-1)}}{c_{n-1}^{(m-1)}(n+2) + n^2 + n - 1}(\alpha_1 + \alpha_2 + \cdots + \alpha_m)$$

because of (85). Hence Theorem 10 is proved.

COROLLARY: *If $A_1 + A_2 + \cdots + A_n$ is not the sum of all positive integers, we have*

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m < 1/c_n^{(m)}.$$

The constants $c_n^{(m)}$ can be determined successively by (83) and (84). We have $c_n^{(1)} = n$,

$$c_n^{(2)} = \frac{n}{2 + 1/(n-1)(n+2)},$$

and so on.

We want to obtain estimates for $c_n^{(m)}$. To this end we prove

THEOREM 11: *Set* $n - m = d$ *for* $n > m$. *For a fixed value of* d *we denote by* z *the greatest integer such that*

$$\log (z-1) + \frac{1}{z+1} \leqq d + \log d + 2. \tag{86}$$

Then we have

$$c_n^{(n-d)} > \frac{n+1}{n-d+\log n - \log d} \qquad \text{for } d+2 \leqq n \leqq z, \tag{87}$$

$$c_n^{(n-d)} > \frac{z+1}{z+2} > \frac{e^{d+1}d+1}{e^{d+1}d+2} \qquad \text{for } n \geqq z. \tag{88}$$

PROOF: First we assume that

$$c_{n-1}^{(m-1)} > \frac{n+1}{n+2}. \tag{89}$$

We maintain that then

$$c_{n-2}^{(m-2)} > \frac{n}{n+1}. \tag{90}$$

For, if $c_{n-2}^{(m-2)} \leqq n/(n+1)$ it follows from (84) that

$$c_{n-1}^{(m-1)} = c_{n-2}^{(m-2)} \leqq \frac{n}{n+1} < \frac{n+1}{n+2}.$$

This contradicts (89). Therefore (90) is proved; we infer by induction

$$c_{n-\mu}^{(m-\mu)} > \frac{n-\mu+2}{n-\mu+3} \qquad \text{for } \mu = 2, 3, \cdots, m-1 \tag{91}$$

if (89) is satisfied. Because of (91) we obtain $c_{n-m+2}^{(2)}, c_{n-m+3}^{(3)}, \cdots, c_n^{(m)}$ successively from (83).

We set $y_n^{(m)} = 1/c_n^{(m)}$. It follows from (83) that

$$ny_n^{(m)} = \frac{n^2+n-1}{n+2}\, y_{n-1}^{(m-1)} + 1 < \frac{n^2}{n+1}\, y_{n-1}^{(m-1)} + 1,$$

hence

$$(n+1)y_n^{(m)} < ny_{n-1}^{(m-1)} + \frac{n+1}{n}.$$

Applying this formula successively we obtain

$$\begin{aligned}(n+1)y_n^{(m)} &< (n-1)y_{n-2}^{(m-2)} + \frac{n}{n-1} + \frac{n+1}{n} < \cdots \\ &< (n-m+2)y_{n-m+1}^{(1)} + \frac{n-m+3}{n-m+2} + \frac{n-m+4}{n-m+3} + \cdots + \frac{n+1}{n} \\ &= \frac{n-m+2}{n-m+1} + \frac{n-m+3}{n-m+2} + \cdots + \frac{n+1}{n} \\ &= m + \sum_{\nu=n-m+1}^{n} \frac{1}{\nu} < m + \int_{n-m}^{n} \frac{dx}{x} = m + \log \frac{n}{n-m}.\end{aligned}$$

Therefore we have

$$c_n^{(m)} > \frac{n+1}{m + \log n - \log (n-m)} \tag{92}$$

if (89) is satisfied.

Next we want to prove that (89) is satisfied if

$$d + 2 \leqq n \leqq z$$

where z is defined by (86). To this end we prove that

$$c_{d+r}^{(r)} > \frac{d+r+1}{r + \log (d+r) - \log d} \geqq \frac{d+r+2}{d+r+3} \tag{93}$$

if the condition

$$d + 2 \leqq d + r \leqq z - 1 \tag{94}$$

is satisfied.

We have

$$c_{d+1}^{(1)} = d + 1 > (d+3)/(d+4),$$

hence it follows from (92) that

$$c_{d+2}^{(2)} > \frac{d+3}{2 + \log (d+2) - \log d} > 1 > \frac{d+4}{d+5}$$

since

$$d + 3 \geqq 4 > 2 + \log 3 \geqq 2 + \log \left(1 + \frac{2}{d}\right) = 2 + \log (d+2) - \log d.$$

Therefore (93) is proved for $r = 2$. We assume it to be proved by induction for $r = 2, 3, \cdots, \rho - 1$ where

$$d + \rho - 1 \leqq z - 2. \tag{95}$$

We then have

$$c_{d+\rho-1}^{(\rho-1)} > \frac{d+\rho}{\rho - 1 + \log (d + \rho - 1) - \log d} \geqq \frac{d+\rho+1}{d+\rho+2}.$$

Hence the condition (89) is satisfied for $n = d + \rho$ and it follows from (92) that

$$c_{d+\rho}^{(\rho)} > \frac{d+\rho+1}{\rho + \log (d+\rho) - \log d}. \tag{96}$$

On the other hand, it follows from (95) that $d + \rho + 1 \leqq z$ and therefore from (86) that

$$\log (d+\rho) + \frac{1}{d+\rho+2} \leqq d + \log d + 2.$$

Consequently we have

$$d + \rho + 2 - \frac{1}{d + \rho + 2} \geqq \rho + \log (d + \rho) - \log d,$$

$$\frac{(d + \rho + 1)(d + \rho + 3)}{d + \rho + 2} \geqq \rho + \log (d + \rho) - \log d,$$

hence by (96)

$$c_{d+\rho}^{(\rho)} > \frac{d + \rho + 1}{\rho + \log (d + \rho) - \log d} \geqq \frac{d + \rho + 2}{d + \rho + 3}.$$

Therefore (93) is proved also for $r = \rho$ and hence by induction for each r satisfying (94). It follows from (93) for $d + r = z - 1$ that (89) is true for $n = z$. This proves (87) because of (92).

Now we have to prove (88). Let s be the greatest integer such that

$$c_{z-1+s}^{(z-1+s-d)} > \frac{z + s + 1}{z + s + 2};$$

then we have $s \geqq 0$ because of (93). It follows similarly as above that

(97) $$c_{z-1+\sigma}^{(z-1+\sigma-d)} > \frac{z + \sigma + 1}{z + \sigma + 2} \geqq \frac{z + 1}{z + 2} \qquad \text{for } \sigma = 0, 1, \cdots, s.$$

Furthermore we have

$$c_{z-1+\sigma}^{(z-1+\sigma-d)} \leqq \frac{z + \sigma + 1}{z + \sigma + 2} \qquad \text{for } \sigma \geqq s + 1,$$

hence by (84)

(98) $$c_{z-1+\sigma}^{(z-1+\sigma-d)} = c_{z-1+\sigma-1}^{(z-1+\sigma-1-d)} = \cdots = c_{z+s}^{z+s-d} \qquad \text{for } \sigma \geqq s + 2.$$

Finally we have to consider the case $n = z + s$. Denoting the density of $A_2 + A_3 + \cdots + A_{z+s}$ by β we obtain from VIII

$$\gamma \geqq \beta + \frac{z + s + 1}{z + s + 2} \alpha_1 \geqq c_{z+s-1}^{(z+s-1-d)} (\alpha_2 + \alpha_3 + \cdots + \alpha_{z+s-d}) + \frac{z + s + 1}{z + s + 2} \alpha_1$$

$$> \frac{z + 1}{z + 2} (\alpha_1 + \alpha_2 + \cdots + \alpha_{z+s-d})$$

because of (97); therefore we have

(99) $$c_{z+s}^{(z+s-d)} > \frac{z + 1}{z + 2}.$$

It follows from (98) and (99) that

(100) $$c_{z-1+\sigma}^{(z-1+\sigma-d)} > \frac{z + 1}{z + 2} \qquad \text{for } \sigma \geqq s + 1.$$

Moreover, the condition (86) implies

$$\log z + \frac{1}{z+2} > d + \log d + 2, \tag{101}$$

$$z > e^{d+1} d.$$

Now (88) follows from (97) and (100). Hence Theorem 11 is proved.

The inequality (88) can be improved a little. Since $d \geqq 1$, it follows from (101) that

$$\log z + \frac{1}{z+2} > 3,$$

$$z \geqq 20, \tag{102}$$

hence we obtain from (101)

$$\log z > d + \log d + \tfrac{43}{22},$$

$$c_n^{(n-d)} > \frac{z+1}{z+2} > \frac{e^{d+43/22}d+1}{e^{d+43/22}d+2} \qquad \text{if } n \geqq z.$$

I. Schur has proved that

$$\gamma \geqq g_n^{(m)}(\alpha_1 + \alpha_2 + \cdots + \alpha_m)$$

where the constants $g_n^{(m)}$ are defined by

$$g_n^{(1)} = n, \qquad g_n^{(m)} = \frac{n}{1 + (n^2-n)^{\frac{1}{2}}/g_{n-1}^{(m-1)}} \qquad \text{for } m > 1. \tag{103}$$

We want to prove that $c_n^{(m)} \geqq g_n^{(m)}$. For $m = 1$ we have $c_n^{(1)} = g_n^{(1)} = n$. Furthermore we have for $n \geqq 2$

$$(n^2-n)(n+2)^2 = n^4 + 3n^3 - 4n > n^4 + 2n^3 - n^2 - 2n + 1 = (n^2+n-1)^2,$$

hence

$$(n^2-n)^{\frac{1}{2}} > \frac{n^2+n-1}{n+2}.$$

Therefore it follows from (103) that we have $c_n^{(m)} \geqq g_n^{(m)}$ for all those $c_n^{(m)}$ which are defined by (83).

From the proof of Theorem 11 we deduce that we have to apply (83) certainly as long as $n \leqq z$ for a fixed value of d. Therefore we may assume now that

$$n \geqq z + 1 > e^{d+1}d + 1.$$

It follows from (88) that it is sufficient to prove that

$$\frac{e^{d+1}d+1}{e^{d+1}d+2} \geqq g_n^{(n-d)} \qquad \text{for } n > e^{d+1}d + 1. \tag{104}$$

From (103) Schur obtained

$$g_n^{(n-d)} = \frac{n^{\frac{1}{2}}}{(d+1)^{-\frac{1}{2}} + (d+2)^{-\frac{1}{2}} + \cdots + n^{-\frac{1}{2}}}. \tag{105}$$

We have

$$\begin{cases} (d+1)^{-\frac{1}{2}} + (d+2)^{-\frac{1}{2}} + \cdots + n^{-\frac{1}{2}} > \displaystyle\int_{d+1}^{n+1} x^{-\frac{1}{2}}\, dx \\ \qquad\qquad = 2(n+1)^{\frac{1}{2}} - 2(d+1)^{\frac{1}{2}} > 2n^{\frac{1}{2}} - 2\,(d+1)^{\frac{1}{2}}. \end{cases} \tag{106}$$

Because of (104), (105), and (106), it is sufficient to prove that

$$\frac{e^{d+1}d+1}{e^{d+1}d+2} > \frac{n^{\frac{1}{2}}}{2n^{\frac{1}{2}} - 2(d+1)^{\frac{1}{2}}}.$$

This is satisfied if

$$e^{d+1}dn^{\frac{1}{2}} > 2(e^{d+1}d+1)(d+1)^{\frac{1}{2}},$$

$$n^{\frac{1}{2}} > 3(d+1)^{\frac{1}{2}},$$

$$n > 9(d+1).$$

This, however, is true for $d = 1$ because of (102), and for $d > 1$ because of (104). Hence we have $c_n^{(m)} \geqq g_n^{(m)}$ and Theorem 10 is sharper than Schur's theorem.

THEOREM 12: *Let ϵ be any positive quantity. For every fixed m and all sufficiently large n we have*

$$\frac{n}{m} - c_n^{(m)} < \epsilon.$$

PROOF: Let m be fixed. We choose n so large that the following conditions are satisfied

$$\log(n-1) + \frac{1}{n+1} < n - m + \log(n-m) + 2 \tag{107}$$

and

$$n > \operatorname{Max}(2m, 2/\epsilon). \tag{108}$$

Then it follows from (107), (86), and (87) that

$$c_n^{(m)} > \frac{n+1}{m + \log n - \log(n-m)},$$

$$\frac{n}{m} - c_n^{(m)} < \frac{n}{m} - \frac{n+1}{m + \log(1 + m/(n-m))}.$$

Because of (108) we have $m/(n-m) < 1$, hence

$$\frac{n}{m} - c_n^{(m)} < \frac{n}{m} - \frac{n+1}{m + m/(n-m)}$$
$$= \frac{n}{m} - \frac{(n+1)(n-m)}{m(n-m+1)} = \frac{1}{n-m+1} < \frac{2}{n} < \epsilon.$$

INSTITUTE FOR ADVANCED STUDY
PRINCETON, N. J.

Annals of Mathematics
Vol. 42, No. 4, October, 1941

ON SOME ASYMPTOTIC FORMULAS IN THE THEORY OF THE "FACTORISATIO NUMERORUM"

By P. Erdös

(Received December 2, 1940)

Let $1 < a_1 \leqq a_2 \leqq \cdots$ be a sequence of integers. Denote by $f(n)$ the number of representations of n as the product of the a's, where two representations are considered equal only if they contain the same factors in the same order. As far as I know the first papers written on the subject are those of L. Kalmár,[1] who proved by using the methods of analytic number theory that if $a_k = k + 1$ then

$$F(n) = \sum_{r=1}^{n} f(r) = -\frac{n^\rho}{\rho\zeta'(\rho)}[1 + o(1)], \tag{1}$$

ρ is defined as the unique positive root of $\zeta(\rho) = 2$. He also gives estimates for the error term.

Another paper on this subject is that of E. Hille.[2] He obtains among others the following results: Let $p_1 < p_2 < \cdots$ be a sequence of primes and $a_1 < a_2 < \cdots$ the sequence of integers composed of these primes, then

$$F(n) = cn^\rho[1 + o(1)], \tag{2}$$

where $\sum_i \frac{1}{a_i^\rho} = 1$, $\rho > 0$. Hille uses the theorem of Wiener and Ikehara.

In the present paper we assume that $\sum \frac{1}{a_i^{1+\epsilon}}$ converges for every ϵ and that the a's are not all powers of a_1, then we prove that

$$F(n) = cn^\rho[1 + o(1)], \tag{3}$$

where $\sum_i \frac{1}{a_i^\rho} = 1$, $\rho > 0$. The proof will be elementary.

First we need 2 Lemmas.

Lemma 1

$$F(n) = \sum_k F\left[\frac{n}{a_k}\right] + 1.^3 \tag{4}$$

Proof. Follows immediately by considering those products in which a_k is the first factor, and summing for a_k.

---

[1] L. Kalmár, Acta Litt ac Scient. Szeged, Tom. 5 (1930) p. 95–107.

[2] E. Hille, Acta Arithmetica Vol. 2 (1937) p. 134–146.

[3] The use of this identity was suggested to me by L. Kalmár.

LEMMA 2.

$$0 < \underline{\lim}\, \frac{F(n)}{n^\rho} \leqq \overline{\lim}\, \frac{F(n)}{n^\rho} < \infty. \tag{5}$$

PROOF. Put $F(n) = c_n n^\rho$. We have from (4)

$$c_n n^\rho < \max_{i \leqq \frac{n}{2}} c_i \sum_{a_k} \frac{n^\rho}{a_k^\rho} + 1,$$

hence

$$c_n < \max_{i \leqq \frac{n}{2}} c_i + \frac{1}{n^\rho}.$$

Thus by induction

$$c_n < 1 + \sum_{2^{m-1} < n} \frac{1}{2^{m\rho}} < \infty,$$

which proves the first half of (5).

The proof of the second half of (5) will be slightly more complicated. Put $F(n) = c'_n(n+1)^\rho$. It suffices to prove that $\underline{\lim}\, c'_n > 0$. From $\left[\frac{n}{a_k}\right] \geqq \frac{n+1}{a_k} - 1$ we obtain by (4)

$$c'_n(n+1)^\rho > \min_{i \leqq \frac{n}{2}} c'_i \sum_{a_k \leqq n} \frac{(n+1)^\rho}{a_k^\rho} = \min_{i \leqq \frac{n}{2}} c'_i(n+1)^\rho\left(1 - \sum_{a_k > n} \frac{1}{a_k^\rho}\right).$$

Thus

$$c'_n > \min_{i \leqq \frac{n}{2}} c'_i \left(1 - \sum_{a_k > n} \frac{1}{a_k^\rho}\right).$$

Hence by induction

$$c'_n > \prod_{2^{m-1} < n} \left(1 - \sum_{a_k > 2^m} \frac{1}{a_k^\rho}\right).$$

The product on the right side (if extended to infinity) converges since

$$\sum_{m=1}^{\infty} \sum_{a_k > 2^m} \frac{1}{a_k^\rho} \leqq \sum_{a_k} \frac{\log a_k}{a_k^\rho} < c \sum \frac{1}{a_k^{1+\epsilon}}$$

converges. This proves $\underline{\lim}\, c'_n > 0$, and completes the proof of Lemma 2.

Now we can prove our theorem. Suppose that (3) does not hold, denote

$$0 < c = \underline{\lim}\, \frac{F(n)}{n^\rho} = \underline{\lim}\, \frac{F(n)}{(n+1)^\rho} < \overline{\lim}\, \frac{F(n)}{n^\rho} = \overline{\lim}\, \frac{F(n)}{(n+1)^\rho} = C < \infty. \tag{6}$$

Let m be sufficiently large and such that $F(m) > (C - \delta)(m+1)^\rho$. Clearly a fixed k exists (depending only on c and C) such that for every x satisfying $m \leqq x \leqq m(1+k)$

$$\frac{F(x)}{(x+1)^\rho} > \frac{C+c}{2}. \tag{7}$$

Now let a_i be the least a which is not a power of a_1. Consider any x satisfying $ma_1 \leqq x \leqq ma_1(1+k)$. We have by (4), (6), (7) and $\left[\frac{x}{a_i}\right] + 1 \geqq \frac{x+1}{a_i}$

$$F(x) > \sum_{a_i \leqq x} F\left[\frac{x}{a_i}\right] > \frac{c+C}{2}\frac{(x+1)^\rho}{a_1^\rho} + c \sum_{a_i > a_1} \frac{(x+1)^\rho}{a_i^\rho} - o(x^\rho). \tag{8}$$

Thus

$$\frac{F(x)}{(x+1)^\rho} > c + \frac{C-c}{2a_1^\rho} - o(1). \tag{9}$$

Similarly we obtain that for the x satisfying $a_1^\alpha a_i^\beta m \leqq x \leqq a_1^\alpha a_i^\beta m(1+k)$

$$\frac{F(x)}{(x+1)^\rho} > c + \delta_{\alpha,\beta}, \tag{10}$$

where $\delta_{\alpha,\beta}$ depends only upon α and β. It is well known that the quotient of two consecutive integers of the form $a_1^\alpha a_i^\beta$ tends to 1. Thus there exists a sequence of integers $A_1 < A_2 < \cdots < A_r$ all of the form $a_1^\alpha a_i^\beta$ and satisfying

$$\frac{A_{i+1}}{A_i} < 1 + k, \qquad i = 1, 2, \cdots r-1 \quad \text{and} \quad A_r > a_1 A_1.$$

Thus by (10) and since the intervals $[A_i m, A_i m(1+k)]$ and $[A_{i+1}m, A_{i+1}m(1+k)]$ overlap we have for $A_1 m \leqq x \leqq a_1 A_1 m$

$$\frac{F(x)}{(x+1)^\rho} > c + \min \delta_{\alpha,\beta} = c + \delta, \tag{11}$$

for sufficiently large m, where δ is fixed and depends only on c and C. Consider now the integers x satisfying $a_1 A_1 m \leqq x \leqq a_1^2 A_1 m_1$ by (4), (6) and (11) we obtain as in (8) and (9)

$$\frac{F(x)}{(x+1)^\rho} > (c+\delta)\frac{1}{a_1^\rho} + c\sum_{a_i > a_1}\frac{1}{a_i^\rho} - o(1) = c + \delta\left(1 - \sum_{a_i > a_1}\frac{1}{a_i^\rho}\right) - o(1).$$

$\left(\text{i.e. } \frac{x}{a_1} \text{ lies in } [A_1 m, A_1 m(1+k)]\right)$. Similarly for the integers satisfying $a_1^2 A_1 m \leqq x \leqq a_1^3 A_1 m$ we have

$$\frac{F(x)}{(x+1)^\rho} > \left[c + \delta\left(1 - \sum_{a_i > a_1}\frac{1}{a_i^\rho}\right)\right]\sum_{a_i \leqq a_1^2}\frac{1}{a_i^\rho} + c\sum_{a_i > a_1^2}\frac{1}{a_i^\rho}$$
$$- o(1) > c + \delta\left(1 - \sum_{a_i > a_1}\frac{1}{a_i^\rho}\right)\left(1 - \sum_{a_i > a_1^2}\frac{1}{a_i^\rho}\right) - o(1).$$

Finally we obtain for $a_1^{k-1}A_1m \leqq x \leqq a_1^kA_1m$ (k fixed, m sufficiently large)

$$\frac{F(x)}{(x+1)^\rho} > c + \delta \prod_{r=1}^{k}\left(1 - \sum_{a_i > a_1^r} \frac{1}{a_i^\rho}\right) - o(1). \tag{12}$$

Denote

$$\prod_{r=1}^{\infty}\left(1 - \sum_{a_i > a_1^r} \frac{1}{a_i^\rho}\right) = \eta.$$

The product converges since $\sum \frac{\log a_i}{a_i^\rho}$ converges. From (12) we have for $A_1m \leqq x \leqq a_1^kA_1m$

$$\frac{F(x)}{(x+1)^\rho} > c + \frac{\delta\eta}{2}. \tag{13}$$

Now choose k so great that

$$\prod_{r>k} \sum_{a_i \leqq a_1^r} \frac{1}{a_i^\rho} > \frac{c + \frac{1}{4}\delta\eta}{c + \frac{1}{2}\delta\eta}. \tag{14}$$

Then from (13) and (4) we have for $A_1a_1^km \leqq x \leqq A_1a_1^{k+1}m$

$$F(x) > \sum_{a_i \leqq a_1^{k+1}} F\left[\frac{x}{a_i}\right] > \left(c + \frac{\delta\eta}{2}\right) \sum_{a_i \leqq a_1^{k+1}} \frac{(x+1)^\rho}{a_i^\rho}.$$

Similarly for any r, in the interval $A_1a_1^rm \leqq x \leqq A_1a_1^{r+1}m$ we have by (14)

$$\frac{F(x)}{(x+1)^\rho} > \left(c + \frac{\delta\eta}{2}\right) \prod_{l>k} \sum_{a_i < a_l} \frac{(x+1)}{\rho} > \frac{c + \delta\eta}{4}.$$

Thus $\lim \frac{F(x)}{(x+1)^\rho} > c$. This contradicts (6) and completes the proof of our theorem.

It is easy to see that in our theorem, we can replace the assumption that $\sum \frac{1}{a_i^{1+\epsilon}}$ converges by the following slightly more general one: There exists a $k > 0$ such that $\sum \frac{1}{a_i^k}$ converges, and $\sum \frac{\log a_i}{a_i^k}$ converges too.

Let $a_k = k + 1$. By using Lemma 2 we can prove that constants c_1 and c_2 exist, $0 < c_2 < c_1 < 1$, such that for infinitely many n

$$f(n) > \frac{n^\rho}{e^{(\log n)^{c_1}}}$$

and that for all $n > n_0$

$$f(n) < \frac{n^{\rho}}{e^{(\log n)^{c_2}}}.^{4}$$

As I shall show in another paper the methods used here yield some asymptotic formulas in the theory of partitions.

THE UNIVERSITY OF PENNSYLVANIA

[4] E. Hille proved that $f(n) > n^{\rho-\epsilon}$ for infinitely many n (ibid).

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

CONCRETE REPRESENTATION OF ABSTRACT (M)-SPACES
(A characterization of the space of continuous functions)

BY SHIZUO KAKUTANI

(Received December 23, 1940)

Introduction

A Banach space is called an abstract (M)-space if it is a linear lattice, and if it satisfies, besides the usual axioms of Banach lattices, the following condition:

$$x \geqq 0, \qquad y \geqq 0 \qquad \textit{imply} \ \| x \vee y \| = \max (\| x \|, \| y \|).$$

The principal purpose of the present paper is to show that every such abstract (M)-space can be concretely represented isometrically and lattice isomorphically by a subspace of the space $C(\Omega)$ of all continuous real valued functions defined on a certain compact Hausdorff space Ω.[1] This may be viewed as a characterization of the subspace of the space of continuous functions, which is closed both in a sense of topology and lattice. Furthermore, this problem has close relations with the problems of concrete representation of abstract (L)-spaces, discussed by the author [9], [11] in connection with the general theory of Markoff processes and ergodic theorems.

In Part I, we shall give the definitions and the exact formulations of fundamental theorems (Theorems 1 and 2). We shall also discuss their relations to other problems concerning the concrete representation of abstract spaces and rings. The proofs of Theorems 1 and 2 will be given in Part II. The main idea is to consider a class of functionals (viz., linear- and lattice-homomorphisms of a given space to the space of real numbers) and to use the notion of weak topology. The last Part III is devoted to the applications of fundamental theorems, and the relations with the theory of abstract (L)-spaces will also be discussed.

The principal results and the general outline of their proofs were previously announced in [10]. I should like to express my hearty thanks to Professors H. F. Bohnenblust, C. Chevalley and J. von Neumann for their kind discussions in the course of this work.[2]

[1] We use the term "compact" as "open coverings can be reduced to finite coverings."

[2] After completing the paper, I noticed that the same problem was discussed by Mark and Selim Krein (On an inner characteristic of the set of all continuous functions defined on a bicompact Hausdorff space, C. R. URSS, 27 (1940), 427–430). In this paper the existence of a unit element is always assumed; Theorem 2 was proved by them, but the case of Theorem 1 was not discussed.

I. Definitions and the Statement of Fundamental Theorems

1. Definition of abstract (M)-spaces

A Banach space (AM) is called an *abstract* (M)*-space*, if there is defined a relation $x \geqq y$ (or equivalently $y \leqq x$) for some pairs of elements $x, y \in (AM)$, and if it satisfies the following conditions ($x, y, z, w \in (AM)$, λ: real number):

(1.1) $\quad x \geqq y, \quad y \geqq x$ *imply* $\quad x = y,$

(1.2) $\quad x \geqq y, \quad y \geqq z$ *imply* $\quad x \geqq z,$

(1.3) $\quad x \geqq y, \quad \lambda \geqq 0$ *imply* $\quad \lambda x \geqq \lambda y,$

(1.4) $\quad x \geqq y$ *implies* $\quad x + z \geqq y + z,$

(1.5) *To any pair of elements* $x, y \in (AM)$, *there exists a maximum* $z = x \vee y$ *such that* $z \geqq x, z \geqq y$ *and* $z' \geqq z$ *for any* z' *with* $z' \geqq x, z' \geqq y$.
(1.6) *To any pair of elements* $x, y \in (AM)$, *there exists a minimum* $w = x \wedge y$ *such that* $w \leqq x, w \leqq y$ *and* $w' \leqq w$ *for any* w' *with* $w' \leqq x, w' \leqq y$.

(1.7) $\quad x_n \geqq y_n$, $x_n \to x$ (*strongly*), $\quad y_n \to y$ (*strongly*) $\quad$ *imply* $x \geqq y$,

(1.8) $\quad x \wedge y = 0$ $\quad$ *implies* $\quad \| x + y \| = \| x - y \|,$

(1.9) $\quad x \geqq 0, \quad y \geqq 0$ *imply* $\quad \| x \vee y \| = \max (\| x \|, \| y \|).$

The conditions (1.1)–(1.6) mean that (AM) is a linear lattice (or a vector lattice). Such a linear space was discussed by G. Birkhoff [1], H. Freudenthal [5] and L. Kantorovitch [12]. The general theory of such linear lattices will be found in G. Birkhoff [2]. Here we shall state some of the fundamental properties of such spaces, which we shall need in the following discussions:

$$(1.10) \qquad \begin{aligned} \lambda \geqq 0 \quad \textit{implies} \quad \lambda(x \vee y) &= \lambda x \vee \lambda y, \\ \lambda(x \wedge y) &= \lambda x \wedge \lambda y, \end{aligned}$$

$$(1.11) \qquad \begin{aligned} (x \vee y) + z &= (x + z) \vee (y + z), \\ (x \wedge y) + z &= (x + z) \wedge (y + z), \end{aligned}$$

$$(1.12) \qquad (x \vee y) + (x \wedge y) = x + y,$$

$$(1.13) \qquad x = x \vee 0 - (-x) \vee 0, \qquad (x \vee 0) \wedge ((-x) \vee 0) = 0,$$

$$(1.14) \qquad (x \vee y) \wedge z = (x \wedge z) \vee (y \wedge z),$$

$$(1.15) \qquad (x \wedge y) \vee z = (x \vee z) \wedge (y \vee z).$$

We shall make use of these properties without any reference to them.

Among all conditions (1.1)–(1.9), the last condition (1.9) is the most important, although, as is easily seen, the condition (1.8) is also indispensable in the following discussions.

It is clear that (1.9) implies:

(1.16) $x \geqq y \geqq 0$ *implies* $\| x \| \geqq \| y \| \geqq 0$,

(1.17) $x \wedge y = 0$ *implies* $\| x \vee y \| = \max (\| x \|, \| y \|)$.

These conditions were introduced by H. F. Bohnenblust [3]. But we do not know whether or not, conversely, the conditions (1.16) and (1.17) together imply (1.9).[2,a]

2. Examples of abstract (*M*)-spaces

The foremost example of abstract (M)-spaces will be given by the space $C(\Omega)$ of all bounded continuous real-valued functions $x(t)$ defined on a Hausdorff space Ω where $\| x \| = \text{l.u.b.}_{t \in \Omega} | x(t) |$ and $x \geqq y$ if and only if $x(t) \geqq y(t)$ for any $t \in \Omega$. If we now take an arbitrary set Ω and consider it as a discrete space, then the space $C(\Omega)$ is nothing but the space $M(\Omega)$ of all bounded real-valued functions $x(t)$ defined on Ω (with the same norm and partial ordering as in the case of $C(\Omega)$).

Moreover, if we consider a subspace $M(\Omega; m)$ of $M(\Omega)$ consisting of all bounded measurable real-valued functions $x(t)$ defined on Ω (where measurability is with respect to a measure $m(E)$ defined *a priori* on Ω), then this is also an example of abstract (M)-spaces. Furthermore, a space of the same kind will be obtained if we neglect some class of subsets of Ω. For example, in the case of $M(\Omega; m)$, if we neglect sets of measure zero of Ω, then the space $M'(\Omega; m)$ thus obtained is again an example of abstract (M)-spaces. It is clear that in this case we have to put $\| x \| = \text{ess. max.}_{t \in \Omega} | x(t) |$[3] and $x \geqq y$ if and only if $x(t) \geqq y(t)$ almost everywhere on Ω. (Two functions which differ from each other only on a set of measure zero are considered as the same element of $M'(\Omega; m)$.)

Another important example of abstract (M)-spaces will be obtained by considering a special subspace of $C(\Omega)$. Indeed, if we consider only those functions $x(t)$ of $C(\Omega)$ which vanish at a given point $t_0 \in \Omega$, then the space $C(\Omega; t_0)$ of all such functions $x(t)$ will be an example. More generally, if we consider the subspace $C(\Omega; t_0, t_0'; \lambda_0)$ of $C(\Omega)$ consisting of all functions $x(t) \in C(\Omega)$ which satisfy the relation:

$$x(t_0) = \lambda_0 x(t_0'), \tag{2.1}$$

where $t_0, t_0' \in \Omega$ and $0 \leqq \lambda_0 \leqq 1$, then this is again an example. Furthermore, it will be easily seen that the spaces of the same kind will be obtained if we replace the point t_0 of the former example by a closed set F, and if we replace the relation (2.1) in the latter example by the system of relations:

$$x(t_\alpha) = \lambda_\alpha x(t_\alpha'), \qquad \alpha \in \mathfrak{M}, \tag{2.2}$$

[2a] See the following paper.

[3] ess. $\max._{t \in \Omega} | x(t) |$ means the greatest lower bound of all $\alpha > 0$ such that $m[t: | x(t) | > \alpha] = 0$.

where t_α, $t'_\alpha \in \Omega$, $0 \leqq \lambda_\alpha \leqq 1$ and $\mathfrak{M}$ is a set of indices α of any power. We shall denote these spaces by $C(\Omega; F)$ and $C(\Omega; t_\alpha, t'_\alpha; \lambda_\alpha; \alpha \in \mathfrak{M})$ respectively.

It is to be noted that, if in the case of $C(\Omega; F)$, the closed set F consists of more than one point, then we can identify all points of F, and that we can thus reduce the space $C(\Omega; F)$ to the case of $C(\Omega; t_0)$. This fact is also true for the spaces $C(\Omega; t_\alpha, t'_\alpha; \lambda_\alpha; \alpha \in \mathfrak{M})$ if there exists a λ_α with $\lambda_\alpha = 1$.

3. Unit element

If there exists an element $\mathbf{1}$ such that

$$\mathbf{1} \geqq 0, \qquad \|\mathbf{1}\| = 1, \tag{3.1}$$

$$\|x\| \leqq 1 \quad \textit{implies} \quad x \leqq \mathbf{1}, \tag{3.2}$$

then $\mathbf{1}$ is called a *unit element*.

It is clear that there exists at most one unit element. But the existence of a unit element is not assumed in abstract (M)-spaces. Indeed, among the examples given in section 2, the spaces $C(\Omega)$, $M(\Omega)$, $M(\Omega; m)$ and $M'(\Omega; m)$ have unit elements (in this case, $x(t) \equiv 1$ is a unit element), while this is not the case for $C(\Omega; t_0)$ and $C(\Omega; t_0, t'_0; \lambda_0)$ if the point t_0 is not isolated in Ω.

Remark 1. The existence of a unit element together with the condition (1.16) will imply (1.9). Indeed, given $x \geqq 0$, $y \geqq 0$, we have, by (3.2), $x \leqq \|x\| \cdot \mathbf{1}$, $y \leqq \|y\| \cdot \mathbf{1}$, and consequently $x \vee y \leqq \max(\|x\|, \|y\|) \cdot \mathbf{1}$, which in turn implies $\|x \vee y\| \leqq \max(\|x\|, \|y\|)$ by (3.1). The inverse relation $\|x \vee y\| \geqq \max(\|x\|, \|y\|)$ is clear from (1.16).

Remark 2. It will be easily seen that $\mathbf{1}$ is also a unit element in the sense of H. Freudenthal [5] (notation: F-unit) which is defined by the following condition:

$$x > 0 \quad \textit{implies} \quad x \wedge \mathbf{1} > 0. \tag{3.3}$$

Indeed, in case $\|x\| \leqq 1$ this follows directly from (3.2), and if $\|x\| > 1$, then $x \wedge \mathbf{1} \geqq \|x\|^{-1} \cdot x \wedge \mathbf{1} = \|x\|^{-1} \cdot x > 0$.

Moreover, an F-unit can exist even if the unit element in our sense does not exist. For example, consider the space $C(\Omega_0; 0)$ or $C(\Omega_0; 0, 1; \frac{1}{2})$, where Ω_0 is the closed interval $0 \leqq t \leqq 1$. These are the spaces of all continuous real-valued functions $x(t)$ defined on $0 \leqq t \leqq 1$ such that $x(0) = 0$ or $x(0) = \frac{1}{2}x(1)$ respectively. It is easy to see that there exists an F-unit in these spaces, while there is no unit in our sense at all. This fact is also true for general spaces $C(\Omega; t_0)$ and $C(\Omega; t_0, t'_0; \lambda_0)$ $(\lambda_0 < 1)$ if the point t_0 in question is not isolated in Ω, and if the topology of Ω satisfies the first countability axiom at $t = t_0$. But in a more general case when the topology of Ω does not satisfy the first countability axiom at $t = t_0$, this is no longer true. Indeed, let us consider the space $D(\Omega)$ of all bounded real-valued functions $x(t)$ defined on Ω such that the set of points at which we have $|x(t)| > \epsilon$ is finite for any $\epsilon > 0$. Then $D(\Omega)$ is an abstract (M)-space if we define the norm and the partial ordering in

the same way as in the case of $C(\Omega)$. It is easy to see that there exists no F-unit in this (M)-space, and it is not without interest to notice that, if we adjoin a new point t_0 to Ω and if we introduce a topology of $\Omega^* = \Omega + (t_0)$ in the following manner:

(3.4) *Every point of* Ω *is isolated,*

(3.5) *Every neighborhood* $U(t_0)$ *of* t_0 *is a set of the form*: $U(t_0) = (t_0) + (\Omega - A)$, *where* A *is an arbitrary finite subset of* Ω, then Ω^* is a compact Hausdorff space and the space $C(\Omega;\, t_0)$ is the space $D(\Omega)$ in question.

4. Fundamental theorems

After the preliminaries given above, we shall now state the fundamental theorems.

THEOREM 1. *For any abstract* (M)*-space* (AM) *(with or without a unit element), there exists a compact Hausdorff space* Ω *and a system of pairs of points* $\{t_\alpha\}$, $\{t'_\alpha\}$ $(t_\alpha \in \Omega,\ \alpha \in \mathfrak{M})$ *and real numbers* $\{\lambda_\alpha\}$ $(0 \leq \lambda_\alpha < 1)$ $(\alpha \in \mathfrak{M})$ *such that* (AM) *is isometric and lattice isomorphic to the space* $C(\Omega;\, t_\alpha,\, t'_\alpha;\, \lambda_\alpha;\, \alpha \in \mathfrak{M})$ *of all bounded continuous real-valued functions* $x(t)$ *which are defined on* Ω *and which satisfy the following relations*:

$$x(t_\alpha) = \lambda_\alpha x(t'_\alpha), \qquad \alpha \in \mathfrak{M}, \tag{4.1}$$

where $\mathfrak{M}$ *is a set of indices* α *whose power can be arbitrarily large.*

This theorem is complicated, and the reason for this is that we do not assume the existence of a unit element. If we assume the existence of a unit element, then it becomes simpler:

THEOREM 2. *For any abstract* (M)*-space* (AM) *with a unit (see section 3), there exists a compact Hausdorff space* Ω *such that* (AM) *is isometric and lattice isomorphic to the space* $C(\Omega)$ *of all bounded continuous real-valued functions* $x(t)$ *defined on* Ω.[4]

Proofs of Theorems 1 and 2 will be given in sections 7 and 11 respectively. We can also prove that the compact Hausdorff space Ω is uniquely determined up to a homeomorphism. This problem will be discussed in section 13 (Theorem 6).

5. Relations to other problems

As is easily observed, this problem has close relations to the problems discussed by I. Gelfand [6], I. Gelfand–A. Kolmogoroff [7], J. von Neumann [13], G. Šilov [16], M. H. Stone [17], [18], and K. Yosida [19]. In all cases discussed by these authors, the Banach space in question was always considered as a ring, and they have obtained a concrete representation of such Banach rings. In our case, however, the space in question is a lattice instead of being a ring, and we can also obtain a concrete representation of such Banach lattices.

Further, it is to be noted that the existence of the unit element (which is

[4] It is not difficult to see that Ω satisfies the second countability axiom of Hausdorff (and hence metrisable) if and only if the given (AM) is separable. It is also obvious that Ω consists only of a finite number of points if and only if (AM) is of finite dimension.

necessary in the case of a ring) is not assumed in Theorem 1. H. F. Bohnenblust [3] has recently discussed the characterization of L_p spaces ($1 \leqq p \leqq \infty$), and the case $p = \infty$ corresponds to our problem. He has only assumed the conditions (1.16) and (1.17), and has obtained a concrete representation of such Banach lattices. But the existence of an F-unit is assumed in his paper, and moreover, besides the separability of the space in question, he assumed that

(5.1) $\quad 0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n \leqq \cdots \leqq \bar{x}$ *implies the existence of* $\sup_n x_n$,

where $x' = \sup_n x_n$ is an element such that $x_n \leqq x'$ for $n = 1, 2, \cdots$, and $x' \leqq x''$ for any x'' with $x_n \leqq x''$ for $n = 1, 2, \cdots$. This condition (5.1), which is also assumed by J. v. Neumann [13] and M. H. Stone [18] (these authors introduced the notions of both ring and lattice), is not satisfied even in the case of the space $C(\Omega_0)$ of all continuous real-valued functions $x(t)$ defined on $\Omega_0 : 0 \leqq t \leqq 1$. Indeed, for example, the sequence of functions $x_n(t)$ ($n = 1, 2, \cdots$), where

$$x_n(t)\begin{cases} = 0, & 0 \leqq t \leqq \frac{1}{2}, \\ = n(t - \frac{1}{2}), & \frac{1}{2} \leqq t \leqq \frac{1}{2} + \dfrac{1}{n}, \\ = 1, & \frac{1}{2} + \dfrac{1}{n} \leqq t \leqq 1, \end{cases}$$

satisfies (5.1) with $\bar{x}(t) \equiv 1$, while there exists no such continuous function $x'(t)$ which corresponds to the element $x' = \sup_n x_n$.

Finally, as will be seen from the papers of J. v. Neumann [13], M. H. Stone [18] and K. Yosida [19], our problem has also a close connection with the theory of spectra for operators.[5]

---

[5] As was remarked by Professor Stone, our problem has close connections with the theory of harmonic functions. Indeed, let D be an arbitrary domain in the Gaussian plane (we do not assume that D is simply connected), and let $H(D)$ be the space of all real valued bounded harmonic functions $u(z)$ defined in D. $H(D)$ is an abstract (M)-space with respect to the norm $\| u \| = \sup_{z \in D} | u(z) |$ and the partial ordering $u \geqq v$: $u(z) \geqq v(z)$ for all $z \in D$. It is to be noticed that $w = u \vee v$ is not equal to the max. $(u(z), v(z))$ in the ordinary sense (maximum taken at each point of D), but $w(z)$ is the smallest harmonic function which satisfies $w(z) \geqq u(z)$ and $w(z) \geqq v(z)$ for all $z \in D$. (The existence of such $w(z)$ follows from the well known principle in the theory of harmonic functions). If, in particular, D is the interior of the unit circle, then every $u(z) \in H(D)$ can be expressed by the Poisson integral:

$$(*) \qquad u(z) = u(re^{i\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{(1 - r^2)x(t)\,dt}{1 - 2r\cos(\theta - t) + r^2},$$

where $x(t)$ is a bounded measurable function defined on the periphery Γ of the unit circle. It is not difficult to see that, if we consider the space $M'(\Gamma, m)$ of all real valued bounded measurable functions $x(t)$ defined on Γ (with respect to the ordinary Lebesgue measure m on Γ; we consider two functions which differ on a set of measure zero to be the same element of $M'(\Gamma, m)$; see section 2), then $M'(\Gamma, m)$ is isometric and lattice isomorphic to the abstract (M)-space $H(D)$, and this isomorphism is given by $(*)$. It is to be noticed that in this case both $H(D)$ and $M'(\Gamma, m)$ satisfy the condition (5.1), and that the representation $(*)$ is isomorphic even with respect to the operation $\sup_n x_n$.

II. Proofs of the Fundamental Theorems

6. Weak topology of Banach spaces

Before going to the proofs of the fundamental theorems, we shall first enter into some elementary considerations on the weak topology of Banach spaces.

Let E be a Banach space and let E^* be its conjugate space, i.e., the Banach space of all bounded linear functionals $f(x)$ defined on E with $\|f\| = \text{l.u.b.}_{\|x\| \leqq 1} |f(x)|$ as its norm.

We shall introduce a weak topology[6] into E^*. For any $f_0 \epsilon E^*$ its weak neighborhood $U(f_0; x_1, x_2, \cdots, x_n; \epsilon)$ is defined as the totality of all $f \epsilon E^*$ such that $|f(x_i) - f_0(x_i)| < \epsilon$ for $i = 1, 2, \cdots, n$, where $\{x_i\}$ $(i = 1, 2, \cdots, n)$ is an arbitrary system of points from E and $\epsilon > 0$ is an arbitrary positive number. The system of all such weak neighborhoods $U(f_0; x_1, x_2, \cdots, x_n; \epsilon)$ determines a Hausdorff topology on E^*, and this topology is called the *weak topology* of E^* (as functionals).

We shall list here some fundamental lemmas on weak topology, which we shall need in the following discussions:[7]

LEMMA 6.1. *For any fixed $x_0 \epsilon E$, $x_0(f) = f(x_0)$ is a continuous function (even a bounded linear functional) defined on E^*, with respect to the weak topology of E^*.*

LEMMA 6.2. *The unit sphere: $\|f\| \leqq 1$ of E^* is compact with respect to the weak topology of E^*.*

LEMMA 6.3. *Let $\Phi(\alpha_1, \alpha_2, \cdots, \alpha_n)$ be a continuous function of n variables $\alpha_1, \alpha_2, \cdots, \alpha_n$, and let $x_1, x_2, \cdots, x_n$ be a given system of points of a Banach space E. Then the totality of all $f \epsilon E^*$ such that $\Phi(f(x_1), f(x_2), \cdots, f(x_n)) = 0$ is a closed subset of E^*.*

7. Outline of the proof of Theorem 1

Let us consider a (real-valued) bounded linear functional $f(x)$ defined on (AM) which satisfies the following conditions:

[6] Weak topology was first introduced by J. von Neumann, Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren, Math. Ann. 102(1930), 370–427. See also: L. Alaoglu, Weak topologies of normed linear spaces, Annals of Math., 41(1940), 252–267; N. Bourbaki, Sur les espaces de Banach, C. R. Paris, 206(1938), 1701–1704; S. Kakutani, Weak topology and the regularity of Banach spaces, Proc. Imp. Acad. Japan, 15(1939), 169–173; A. E. Taylor, Weak topologies of Banach spaces, Proc. Nat. Acad. Sci. U. S. A., 25(1939), 438–440; A. Tychonoff, Über die topologische Erweiterungen von Räumen, Math. Ann., 102(1930), 544–561; A. Tychonoff, Über einen Funktionenraum, Math. Ann., 111(1935,) 762–766.

[7] Lemmas 6.1 and 6.3 follow directly from the definition of weak topology. Lemma 6.2 is a consequence of Tychonoff's theorem that a topological product of an arbitrary (finite or infinite, not necessarily countable) number of closed intervals is compact. (See the papers of A. Tychonoff cited in footnote 6). A simpler proof of this theorem may be found in the following papers, where the result is extended to the case of the topological product of an arbitrary number of compact spaces: J. W. Alexander, Ordered sets, complexes and the problem of compactification, Proc. Nat. Acad. Sci. U. S. A., 25(1939), 296–298; H. Wallman, Lattices and topological spaces, Annals of Math., 39(1938), 112–126. See also E. Čech [4].

(7.1) $$\|f\| = 1,$$

(7.2) $$f(x + y) = f(x) + f(y),$$

(7.3) $$f(\lambda x) = \lambda f(x), \qquad \lambda \gtreqless 0,$$

(7.4) $$x \geqq 0 \quad \textit{implies} \quad f(x) \geqq 0,$$

(7.5) $$x \wedge y = 0 \quad \textit{implies} \quad f(x) = 0 \quad \text{or} \quad f(y) = 0,$$

(7.6) $$f(x \vee y) = \max (f(x), f(y)),$$

(7.7) $$f(x \wedge y) = \min (f(x), f(y)).$$

These conditions are clearly not mutually independent. But it is not necessary to discuss their independence. We shall only notice that the conditions (7.6) and (7.7) follow directly from the other conditions. This fact is needed in the following.

If we consider the space of real numbers as an abstract (M)-space (this is indeed the simplest non-trivial example of abstract (M)-spaces), then these conditions mean that $x \to f(x)$ is a continuous homomorphic mapping of a given space (AM) onto the space of real numbers. (For the general principle of the proofs of such theorems see [10].) Indeed, (7.1) implies that $x \to f(x)$ is continuous; (7.2) and (7.3) mean that $x \to f(x)$ is a linear homomorphism; and lastly, (7.4), (7.5), (7.6) and (7.7) mean that $x \to f(x)$ is a lattice homomorphism.

Let us now consider the totality Ω' of all bounded linear functionals $f(x)$ defined on (AM) which satisfy the conditions (7.1)–(7.7). Ω' is a bounded set contained in the unit sphere of the conjugate space $(AM)^*$ of (AM). Consequently, by Lemma 6.2, the closure $\Omega = \overline{\Omega'}$ of Ω' with respect to the weak topology of $(AM)^*$ is compact, and it is easy to see that every $f \in \Omega$ satisfies the condition (7.2)–(7.7). This is a direct consequence of Lemma 6.3. But, in general, the condition (7.1) is not necessarily satisfied, and we can only say that we have $\|f\| \leqq 1$ for any $f \in \Omega$. If there exists, indeed, an element $f' \in \Omega$ with $\|f\| = \lambda < 1$, then we have $f = \lambda f'$ for some $f' \in \Omega'$. Let us denote all such relations which exist among the elements of Ω by:

(7.8) $$f_\alpha = \lambda_\alpha f'_\alpha, \qquad \alpha \in \mathfrak{M},$$

where $f_\alpha \in \Omega$, $f'_\alpha \in \Omega'$, $0 \leqq \lambda_\alpha < 1$ and $\mathfrak{M}$ is a set of indices α which can be finite, denumerably infinite or even non-denumerable.

We shall next consider the functional $x(f) = f(x)$ defined on Ω (x being fixed). For each $x_0 \in (AM)$, by Lemma 6.1, $x_0(f)$ is a bounded continuous real-valued function defined on Ω, and it is clear that the conditions:

(7.9) $$x(f_\alpha) = \lambda_\alpha x(f'_\alpha), \qquad \alpha \in \mathfrak{M},$$

are all satisfied. Consequently, for each $x_0 \in (AM)$, the function $x_0(f)$ can be considered as an element of $C(\Omega; f_\alpha, f'_\alpha; \lambda_\alpha; \alpha \in \mathfrak{M})$ and the conditions (7.2)–(7.7) imply that $x \to x(f)$ is a continuous (linear and lattice) homomorphism of (AM) into the space $C(\Omega; f_\alpha, f'_\alpha; \lambda_\alpha; \alpha \in \mathfrak{M})$.

Thus, all that we have to do in order to prove Theorem 1, is to show that this mapping $x \to x(f)$ is an isometric mapping of (AM) onto $C(\Omega; f_\alpha, f'_\alpha; \lambda_\alpha; \alpha \in \mathfrak{M})$. It will be easily seen that we have, for this purpose, only to prove the following lemmas:

Lemma 7.1. *For any $x_0 \in (AM)$, there exists an $f_0 \in \Omega$ such that $| f_0(x_0) | = \| x_0 \|$.*

Lemma 7.2. *For any continuous real-valued function $X_0(f) \in C(\Omega; f_\alpha, f'_\alpha; \lambda_\alpha; \alpha \in \mathfrak{M})$, there exists an $x_0 \in (AM)$ such that $f(x_0) = X_0(f)$ for any $f \in \Omega$.*

These are the fundamental lemmas, to whose proofs the remaining part of Part II is devoted. The final proofs of these lemmas will be given in sections 9 and 10 respectively.

8. Lemmas on abstract (M)-spaces

Let $x_0 \in (AM)$ and let $f(x)$ be a bounded linear functional defined on (AM).

Lemma 8.1.[8] *$x_0 \geqq 0$, $\| x_0 \| = 1$, $f(x_0) = 1$, $\| f \| = 1$ imply $f(y) \geqq 0$ for any $y \geqq 0$; i.e., $f(x)$ is a positive linear functional.*

Proof. Without loss of generality, we may assume $\| y \| = 1$. Then, $0 \leqq (x_0 - y) \vee 0 \leqq x_0$ implies $\| (x_0 - y) \vee 0 \| \leqq \| x_0 \| = 1$, and $0 \leqq (y - x_0) \vee 0 \leqq y$ implies $\| (y - x_0) \vee 0 \| \leqq \| y \| = 1$. Hence, $\| x_0 - y \| = \| (x_0 - y) \vee 0 - (y - x_0) \vee 0 \| = \| (x_0 - y) \vee 0 + (y - x_0) \vee 0 \| = \max (\| (x_0 - y) \vee 0 \|, \| (y - x_0) \vee 0 \|) \leqq 1$, and consequently $1 - f(y) = f(x_0) - f(y) = f(x_0 - y) \leqq \| f \| \cdot \| x_0 - y \| \leqq 1$, or equivalently $f(y) \geqq 0$.

Lemma 8.2. *$x_0 \geqq 0$, $\| x_0 \| = 1$, $f(x_0) = 1$, $\| f \| = 1$, $y \geqq 0$, $\| x_0 \wedge y \| < 1$, imply $f(y) = f(x_0 \wedge y) < 1$. In particular, $x_0 \wedge y = 0$ implies $f(y) = 0$.*

Proof. Put $z = (1 - \lambda)(x_0 \wedge y) + \lambda y$, where $0 < \lambda < 1$. Then $x_0 \wedge y \leqq z \leqq y$ and, for sufficiently small λ, we have $\| z \| \leqq (1 - \lambda) \| x_0 \wedge y \| + \lambda \| y \| < 1$. Hence, $x_0 \wedge z = x_0 \wedge (y \wedge z) = (x_0 \wedge y) \wedge z = x_0 \wedge y$, and, since by lemma 8.1 $f(x)$ is a positive linear functional, $1 = f(x_0) \leqq f(x_0 \vee z) \leqq \| f \| \cdot \| x_0 \vee z \| = \| f \| \cdot \max (\| x_0 \|, \| z \|) \leqq 1$, or $f(x_0 \vee z) = 1$. Consequently, $f(x_0 \wedge y) = f(x_0 \wedge z) = f(x_0 + z - x_0 \vee z) = f(x_0) + f(z) - f(x_0 \vee z) = 1 + f(z) - 1 = f(z)$, i.e., $f(x_0 \wedge y) = f(z) = (1 - \lambda) f(x_0 \wedge y) + \lambda f(y)$. Since $\lambda > 0$, we must have $f(x_0 \wedge y) = f(y)$.

Lemma 8.3. *For any $x_0 \geqq 0$ with $\| x_0 \| = 1$, and for any system $\Sigma = \{x_1, x_2, \cdots, x_n\}$ of positive elements of (AM), there exists a $y_0 \in (AM)$ and another system $\{y_1, y_2, \cdots, y_n\}$ of positive elements of (AM) such that*

$$(8.1) \qquad 0 \leqq y_0 \leqq x_0, \qquad \| y_0 \| = 1,$$

$$(8.2) \qquad \| y_0 \wedge \lambda y_i \| < 1 \quad \text{for any} \quad \lambda > 0 \quad \text{and for} \quad i = 1, 2, \cdots, n,$$

$$(8.3) \qquad x_i \wedge x_j = 0 \quad \text{implies} \quad x_i = y_i \quad \text{or} \quad x_j = y_j.$$

Proof. We shall define z_i and y_i ($i = 1, 2, \cdots, n$) by mathematical induction. Put $z_0 = x_0$, and let us assume that $z_{i-1} \geqq 0$ with $\| z_{i-1} \| = 1$ and $y_{i-1} \geqq 0$

[8] Lemmas 8.1 and 8.2 are due to Professor Bohnenblust. By virtue of these lemmas, the proof of Lemma 7.1 became simpler. The original proof of the author was complicated.

are already defined. Then, if there exists a $\lambda_i > 0$ such that $\| z_{i-1} \wedge \lambda_i x_i \| = 1$, put $z_i = z_{i-1} \wedge \lambda_i x_i$ and $y_i = 0$. In other cases (i.e., if we have $\| z_{i-1} \wedge \lambda x_i \| < 1$ for any $\lambda > 0$), put $z_i = z_{i-1}$ and $y_i = x_i$. Thus z_i and y_i are determined for $i = 1, 2, \cdots, n$. If we now put $y_0 = z_n$, then this y_0 and the system $\{y_1, y_2, \cdots, y_n\}$ thus obtained are the required ones.

Indeed, (8.1) is clear. (8.2) is also clear, since $y_i > 0$ implies $y_i = x_i$ and consequently $\| y_0 \wedge \lambda y_i \| = \| y_0 \wedge \lambda x_i \| \leqq \| z_{i-1} \wedge \lambda x_i \| < 1$ for any $\lambda > 0$. In order to prove (8.3), assume $x_i \wedge x_j = 0$ and $i < j$. Then $x_i \neq y_i$ implies $z_i = z_{i-1} \wedge \lambda_i x_i$ for some $\lambda_i > 0$, and consequently $0 \leqq z_{j-1} \wedge \lambda x_j \leqq z_i \wedge \lambda x_j = (z_{i-1} \wedge \lambda_i x_i) \wedge \lambda x_j \leqq \lambda_i x_i \wedge \lambda x_j = 0$ for any $\lambda > 0$, which in turn implies $x_j = y_j$.

LEMMA 8.4. *For any $x_0 \geqq 0$ with $\| x_0 \| = 1$ and for any system $\Sigma = \{x_1, x_2, \cdots, x_n\}$ of positive elements of (AM), there exists a bounded linear functional $f_\Sigma(x)$ defined on (AM) such that*

$$(8.4) \qquad f_\Sigma(x_0) = 1, \qquad \| f_\Sigma \| = 1,$$

$$(8.5) \qquad f_\Sigma(y) \geqq 0 \quad \textit{for any} \quad y \geqq 0,$$

$$(8.6) \qquad x_i \wedge x_j = 0 \quad \textit{implies} \quad f_\Sigma(x_i) = 0 \quad \textit{or} \quad f_\Sigma(x_j) = 0.$$

PROOF. Let y_0 and a system $\{y_1, y_2, \cdots, y_n\}$ be defined as in Lemma 8.3. If we consider a bounded linear functional $f_\Sigma(x)$ which satisfies

$$(8.7) \qquad f_\Sigma(y_0) = 1, \qquad \| f_\Sigma \| = 1$$

(the existence of such $f_\Sigma(x)$ is an easy consequence of the Hahn-Banach extension theorem), then this $f_\Sigma(x)$ is the required one. Indeed, (8.5) is clear by Lemma 8.1, and consequently (8.4) becomes also clear since $1 = f_\Sigma(y_0) \leqq f_\Sigma(x_0) \leqq \| f_\Sigma \| \cdot \| x_0 \| = 1$. In order to prove (8.6) we observe that, by Lemma 8.3, we have either $x_i = y_i$ or $x_j = y_j$. Let us assume $x_i = y_i$. Then, by (8.2) we have $\| y_0 \wedge \lambda x_i \| < 1$ for any $\lambda > 0$. Consequently, Lemma 8.2 implies $f_\Sigma(x_i) = \lambda^{-1} f_\Sigma(\lambda x_i) < \lambda^{-1}$, and since $\lambda > 0$ is arbitrary, we must have $f_\Sigma(x_i) = 0$. Since the case $x_j = y_j$ can be treated in the same way, we have proved (8.6), and the proof of Lemma 8.4 is completed.

9. Proof of Lemma 7.1

First we observe that we may assume $x_0 \geqq 0$ and $\| x_0 \| = 1$. This follows easily from the fact that we have $\| x_0 \| = \| x_0 \vee 0 - (-x_0) \vee 0 \| = \| x_0 \vee 0 + (-x_0) \vee 0 \| = \max (\| x_0 \vee 0 \|, \|(-x_0) \vee 0 \|)$. Indeed, it is clear that we may assume $\| x_0 \| = 1$, and in this case we have either $\| x_0 \vee 0 \| = 1$ or $\|(-x_0) \vee 0 \| = 1$. Let us assume $\| x_0 \vee 0 \| = 1$. If we now assume that Lemma 7.1 is proved for $x_0 \geqq 0$ with $\| x_0 \| = 1$, then there exists a bounded linear functional $f_0 \in \Omega$ such that $f_0(x_0 \vee 0) = 1$ and $\| f_0 \| = 1$. Since $(x_0 \vee 0) \wedge ((-x_0) \vee 0) = 0$, we must have $f_0((-x_0) \vee 0) = 0$ by Lemma 8.2, and consequently $f_0(x_0) = f_0(x_0 \vee 0) - f_0((-x_0) \vee 0) = f_0(x_0 \vee 0)$. Since the case $\|(-x_0) \vee 0 \| = 1$ can be treated analogously, we have to prove Lemma 7.1 only for the case $x_0 \geqq 0$, $\| x_0 \| = 1$.

Let now $x_0 \geqq 0$ be a positive element with $\| x_0 \| = 1$. Then, for any system $\Sigma = \{x_1, x_2, \cdots, x_n\}$ of positive elements of (AM), there exists by Lemma 8.4, a bounded linear functional $f_\Sigma(x)$ which satisfies the conditions (8.4), (8.5) and (8.6). Let Ω_Σ be the set of all such functionals $f_\Sigma(x)$. (Ω_Σ is not necessarily a subset of Ω.) Then Ω_Σ is a non-empty compact set with respect to the weak topology of $(AM)^*$. (See Lemmas 6.2 and 6.3.) Moreover, the family $\{\Omega_\Sigma\}$ of all such Ω_Σ for all choice of Σ has a finite intersection property. This follows directly from the fact that we have $\prod_{i=1}^{n} \Omega_{\Sigma_i} \supseteq \Omega_{\Sigma_1 + \Sigma_2 + \cdots + \Sigma_n} \neq 0$ for any Σ_i ($i = 1, 2, \cdots, n$). Hence there exists, by a well known argument concerning compact sets, a bounded linear functional $f_0(x)$ which belongs to all Ω_Σ. We shall show that this $f_0(x)$ is the required one. Indeed, it is easy to see that $f_0(x)$ has the properties (7.1)–(7.5) (see Lemmas 8.1, 8.4. (7.1) follows from the fact that $f_0(x) = 1$ and $\| f_0 \| \leqq 1$), and since both (7.6) and (7.7) are the direct consequences of (7.1)–(7.5), the proof of Lemma 7.1 is completed.

10. Proof of Lemma 7.2[9]

Let $X_0(f)$ be an arbitrary bounded continuous real-valued function of $C(\Omega; f_\alpha, f_\alpha'; \lambda_\alpha; \alpha \in \mathfrak{M})$. We shall first prove that for any pair of elements f_0 and g_0 of Ω there exists an element $y_0 \in (AM)$ such that

$$(10.1) \qquad f_0(y_0) = X_0(f_0), \qquad g_0(y_0) = X_0(g_0).$$

Indeed, this is clear if there exists an $\alpha \in \mathfrak{M}$ such that $f_0 = f_\alpha$, $g_0 = f_\alpha'$ (or $f_0 = f_\alpha'$, $g_0 = f_\alpha$). In other cases, there exists a pair of points y_0', $y_0'' \in (AM)$ such that

$$\delta = f_0(y_0')g_0(y_0'') - f_0(y_0'')g_0(y_0') \neq 0.$$

Hence, if we put $\alpha = \delta^{-1}(g_0(y_0'')X_0(f_0) - f_0(y_0'')X_0(g_0))$, $\beta = \delta^{-1}(g_0(y_0')X_0(f_0) - f_0(y_0')X_0(g_0))$, then (10.1) is clearly satisfied by $y_0 = \alpha y_0' - \beta y_0''$.

Thus it is proved that there exists, for any $f_0, g_0 \in \Omega$, a point $y_0 \in (AM)$ such that (10.1) is true. Consequently, for any $\epsilon > 0$, there exists a neighborhood $U(f_0)$ of f_0 such that $f(y_0) > X_0(f) - \epsilon$ for any $f \in U(f_0)$. Let us now consider f_0 as a variable ($g_0 \in \Omega$ and $\epsilon > 0$ being fixed). Then there exists, by the compactness of Ω, a finite number of elements $f_i \in \Omega$ ($i = 1, 2, \cdots, m$), their corresponding points $y_i \in (AM)$ ($i = 1, 2, \cdots, m$) and a system of neighborhoods $U(f_i)$ ($i = 1, 2, \cdots, m$) such that $g_0(y_i) = X_0(g_0)$ ($i = 1, 2, \cdots, m$), $\Omega \subseteq \sum_{i=1}^{m} U(f_i)$ and $f(y_i) > X_0(f) - \epsilon$ for any $f \in U(f_i)$. Consequently, if we put $z_0 = y_1 \vee y_2 \vee \cdots \vee y_m$, then we have $g_0(z_0) = \max_{1 \leqq i \leqq m} g_0(y_i) = X_0(g_0)$ and $f(z_0) = \max_{1 \leqq i \leqq m} f(y_i) > X_0(f) - \epsilon$ for any $f \in \Omega$.

Thus we have seen that there exists, for any $g_0 \in \Omega$ and for any $\epsilon > 0$, a point $z_0 \in (AM)$ such that $g_0(z_0) = X_0(y_0)$ and $f(z_0) > X_0(f) - \epsilon$ for any $f \in \Omega$. Clearly, there exists a neighborhood $V(g_0)$ of g_0 such that $f(z_0) < X_0(f) + \epsilon$ for any $f \in Vg_0)$. Let us now consider g_0 as a variable ($\epsilon > 0$ being fixed). Then there

[9] The proof of Lemma 7.2 became simpler by a suggestion of Professor Chevalley.

exist, again by the compactness of Ω, a finite number of elements $g_j \in \Omega (j = 1, 2, \cdots, n)$, their corresponding points $z_j \in (AM)$ $(j = 1, 2, \cdots, n)$ and a system of neighborhoods $V(g_j)$ $(j = 1, 2, \cdots, n)$ such that $\Omega \subseteq \sum_{j=1}^{n} V(g_j), f(z_j) > X_0(f) - \epsilon$ for any $f \in \Omega$ and $f(z_j) < X_0(f) + \epsilon$ for any $f \in V(g_j)$. Consequently, if we put $x_\epsilon = z_1 \wedge z_2 \wedge \cdots \wedge z_n$, then we have $X_0(f) - \epsilon < f(x_\epsilon) = \min_{1 \leqq j \leqq n} f(z_j) < X_0(f) + \epsilon$ for any $f \in \Omega$.

Thus we have proved that there exists, for any $\epsilon > 0$, a point $x_\epsilon \in (AM)$ such that $| f(x_\epsilon) - X_0(f) | < \epsilon$ for any $f \in \Omega$. Let us now consider a sequence $\{x_{\epsilon_n}\}$ $(n = 1, 2, \cdots)$, where $\epsilon_n > 0$, $\epsilon_n \to 0$. Then we have $| f(x_{\epsilon_m} - x_{\epsilon_n}) | < | f(x_{\epsilon_m}) - X_0(f) | + | f(x_{\epsilon_n}) - X_0(f) | < \epsilon_m + \epsilon_n$ for any $f \in \Omega$ and for any m and n. Consequently, by Lemma 7.1, we must have $\| x_{\epsilon_m} - x_{\epsilon_n} \| < \epsilon_m + \epsilon_n$ for any m and n. Thus $\{x_{\epsilon_n}\}$ $(n = 1, 2, \cdots)$ is a fundamental sequence in (AM), and it is clear that $x_0 = \lim_{n\to\infty} x_{\epsilon_n}$ will satisfy $f(x_0) = X_0(f)$ for any $f \in \Omega$. The proof of Lemma 7.1 and consequently the proof of Theorem 1 is completed.

REMARK. It will be easily seen that we can also state Lemma 7.2 in the following form:

THEOREM 3. *Let Ω be a compact Hausdorff space, and let E be a linear subspace of the space $C(\Omega)$ of all bounded continuous real-valued functions defined on Ω, which is closed in $C(\Omega)$ both in the sense of topology and in the sense of lattice. (E is closed in $C(\Omega)$ in the sense of lattice if $x(t), y(t) \in E$ imply* max $(x(t), y(t))$, min $(x(t), y(t)) \in E$.) *If we now denote by $\{t_\alpha, t'_\alpha; \lambda_\alpha; \alpha \in \mathfrak{M}\}$ (where $t_\alpha \in \Omega$, $t'_\alpha \in \Omega$, $0 \leqq \lambda_\alpha \leqq 1$, $(\alpha \in \mathfrak{M})$, and $\mathfrak{M}$ is a set of indices α whose power can be arbitrarily large) the system of all pairs t_α, t'_α, λ_α such that*

$$x(t_\alpha) = \lambda_\alpha x(t'_\alpha), \tag{10.2}$$

for all $x(t) \in E$, then conversely, every function $x(t) \in C(\Omega)$ which satisfies (10.2) *for any $\alpha \in \mathfrak{M}$ belongs to E; i.e., E is the subspace $C(\Omega; t_\alpha, t'_\alpha; \lambda_\alpha; \alpha \in \mathfrak{M})$.*

11. Proof of Theorem 2

LEMMA 11.1. *For every positive bounded linear functional $f(x)$ defined on an abstract (M)-space with a unit element* **1**, *we have $\| f \| = f(\mathbf{1})$.*

PROOF. *Let* $\epsilon > 0$ be an arbitrary positive number. Then there exists an $x \in (AM)$ such that $\| x \| = 1$ and $f(x) > \| f \| - \epsilon$. Since $\| x \| = 1$ implies $x \leqq \mathbf{1}$ by definition, we have, by the positiveness of $f(x)$, $\| f \| - \epsilon < f(x) \leqq f(\mathbf{1}) \leqq \| f \| \cdot \| \mathbf{1} \| = \| f \|$. Since $\epsilon > 0$ is arbitrary, we must have $\| f \| = f(\mathbf{1})$.

Now we shall proceed to the proof of Theorem 2, which is quite easy. We have only to show that, if there exists a unit element **1**, in the given abstract (M)-space (AM), then there exist no relations of the form (7.8). Indeed, we can prove more precisely that, under the same assumption, we have $\Omega = \Omega'$ i.e., Ω' itself is compact with respect to the weak topology of $(AM)^*$.

By Lemma 11.1, every bounded linear functional $f(x) \in \Omega'$ satisfies $f(\mathbf{1}) = 1$. Hence, by Lemma 6.3, every bounded linear functional $f(x) \in \Omega = \overline{\Omega'}$ must also satisfy $f(\mathbf{1}) = 1$ and this in turn implies (again by Lemma 11.1) $\| f \| = 1$ for

any $f(x) \epsilon \Omega$. Consequently, by the definition of Ω, we must have $\Omega = \Omega'$. The proof of Theorem 2 is completed.

We can also state the following

THEOREM 4. *Let Ω be a compact Hausdorff space, and let E be a linear subspace of the space $C(\Omega)$ of all bounded continuous real-valued functions defined on Ω which is closed in $C(\Omega)$ both in the sense of topology and lattice (see Theorem 3). If, moreover, E contains a constant function, and if there exists for any $t_1, t_2 \epsilon \Omega$, $t_1 \neq t_2$, a continuous function $x(t) \epsilon E$ such that $x(t_1) \neq x(t_2)$, then E must be identical with the entire space $C(\Omega)$.*

III. APPLICATIONS

12. Theorem of E. Čech

A Hausdorff space Ω is called *completely regular* if there exists, for any point $t_0 \epsilon \Omega$ and for any neighborhood $U(t_0)$ of t_0, a continuous real-valued function $x(t)$ defined on Ω such that $x(t_0) = 1$ and $x(t) = 0$ for any $t \bar{\epsilon} U(t_0)$.

THEOREM 5. *For any completely regular Hausdorff space Ω', there exists a compact Hausdorff space Ω which contains Ω' as a dense subset and such that every bounded continuous real-valued function defined on Ω' can be uniquely extended to a continuous function defined on Ω.*

This theorem is due to E. Čech [4].

PROOF. Consider the space $C(\Omega')$ of all bounded continuous real-valued functions $x(t)$ defined on Ω'. $C(\Omega')$ is an abstract (M)-space with respect to the norm: $\| x \| = \text{l.u.b.}_{t \epsilon \Omega'} | x(t) |$ and the partial ordering $x \geqq y$: if and only if $x(t) \geqq y(t)$ for all $t \epsilon \Omega'$. Moreover, $x(t) \equiv 1$ is a unit element of $C(\Omega')$. Consequently, by Theorem 2, there exists a compact Hausdorff space Ω such that $C(\Omega')$ is isometric and lattice isomorphic to the space $C(\Omega)$ of all bounded continuous real-valued functions $x(f)$ defined on Ω. (We denote elements of Ω' and Ω by t and f respectively.)

We shall show that Ω' can be homeomorphically embedded into Ω. To prove this, let t_0 be an arbitrary point of Ω', and put $f_0(x) = x(t_0)$ for all $x(t) \epsilon C(\Omega')$. Then $f_0(x)$ is a bounded linear functional defined on $C(\Omega')$ and clearly satisfies the conditions (7.1)–(7.7) stated in section 7. Hence $f_0(x)$ can be considered as a point of Ω. Thus the mapping $t_0 \rightarrow \varphi(t_0) = f_0$ of Ω' onto $\varphi(\Omega') \subseteq \Omega$ is defined, and all that we have to do is to prove that $\varphi(\Omega')$ is dense in Ω and that the mapping $\Omega' \leftrightarrow \varphi(\Omega')$ is a homeomorphism.

In the first place, it is clear that this mapping $t \rightarrow \varphi(t) = f$ is continuous. Indeed, let $U(f_0; x_1, x_2, \cdots, x_n; \epsilon)$ be an arbitrary neighborhood of a point $f_0 = \varphi(t_0) \epsilon \varphi(\Omega')$, where $x_i \epsilon C(\Omega')$ for $i = 1, 2, \cdots, n$, and $\epsilon > 0$. (See the definition of weak topology given in section 6.) Since every $x_i(t)$ is continuous on Ω', there exists a neighborhood $U_i(t_0)$ of t_0 such that $t \epsilon U_i(t_0)$ implies $| x_i(t) - x_i(t_0) | < \epsilon$. Consequently, if we put $U(t_0) = \prod_{i=1}^{n} U_i(t_0)$, then $t \epsilon U(t_0)$ implies $| x_i(t) - x_i(t_0) | < \epsilon$ for $i = 1, 2, \cdots, n$, or equivalently $f = \varphi(t) \epsilon U(f_0; x_1, x_2, \cdots, x_n; \epsilon)$, which means that the mapping $t \rightarrow \varphi(t) = f$ is continuous at $t = t_0$.

Next, it is clear that the mapping $t \to \varphi(t) = f$ is one-to-one. For, by the complete regularity of Ω', there exists, for any two different points $t_0, t_0' \in \Omega'$, $t_0 \neq t_0'$, a bounded continuous real-valued function $x_0(t)$ such that $x_0(t_0) \neq x_0(t_0')$. To prove that the inverse mapping $f \to t = \varphi^{-1}(f)$ is also continuous, let $U(t_0)$ be an arbitrary neighborhood of a point $t_0 \in \Omega'$. Then, again by the complete regularity of Ω', there exists a bounded continuous real-valued function $x_0(t)$ such that $x_0(t_0) = 1$ and $x_0(t) = 0$ for any $t \bar{\in} U(t_0)$. Consequently, every $t \in \Omega'$ which satisfies $\varphi(t) \in U(f_0; x_0; 1)$, where $f_0 = \varphi(t_0)$, must necessarily be inside of $U(t_0)$ (see the definition of weak topology given in section 6), and this fact means that $f \to t = \varphi^{-1}(f)$ is continuous at $f = f_0$.

Thus we have proved that $\Omega' \leftrightarrow \varphi(\Omega') \subseteq \Omega$ is a homeomorphic mapping. It is easy to see that the adjoint mapping $x \leftrightarrow \bar{\varphi}(x) = \bar{x}; \bar{x}(f) = x(\varphi(t))$ is an isometric and lattice-isomorphic mapping between $C(\Omega')$ and $C(\Omega)$. In order to prove that $\varphi(\Omega')$ is dense in Ω, let us assume that the closure $\overline{\varphi(\Omega')}$ of $\varphi(\Omega')$ is a proper subset of Ω. Then there exists a continuous function $x_0(f)$ which is zero at every point of $\varphi(\Omega')$ without vanishing identically on Ω. This is, however, a contradiction, since on the one hand $x_0(f)$ must correspond to a function $x_0(t)$ which is identically zero on Ω' (by the mapping $\bar{\varphi}$), while on the other hand $x_0(f)$ must correspond to a function which is not identically zero on Ω' (by the isometric and lattice-isomorphic mapping of $C(\Omega')$ and $C(\Omega)$).

Thus we have seen that $\varphi(\Omega')$ is dense in Ω. Since the possibility of extension and the uniqueness of extension are both clear from our construction, the proof of Theorem 5 is completed.

REMARK 1. It is to be noted that in the proof given above we have made no essential use of Theorem 2. Indeed, we have only to consider each point $t_0 \in \Omega'$ as a bounded linear functional $f_0(x) = x(t_0)$ defined on $C(\Omega')$. The mapping $t_0 \to \varphi(t_0) = f_0$ thus obtained is a homeomorphic embedding of Ω' into the conjugate space $C^*(\Omega')$ of $C(\Omega')$. This can be proved in exactly the same way as above, and the closure $\Omega = \overline{\varphi(\Omega')}$ of the image of $\varphi(\Omega')$ of Ω' with respect to the weak topology of $C^*(\Omega')$ is the required compact set. It is easy to see that the extension of bounded continuous functions is possible and that this extension is unique.[10]

---

[10] Let G be a topological group and let $B(G)$ be the space of all real valued bounded continuous almost periodic functions $x(t)$ defined on G. $B(G)$ is a Banach space with respect to the norm $\|x\| = \sup_{t\in G} |x(t)|$. It is easy to see that $B(G)$ is an abstract (M)-space with a unit with respect to the ordinary partial ordering $x \geqq y$: $x(t) \geqq y(t)$ for all $t \in G$. (In this case $z = x \vee y$ is exactly the max. $(x(t), y(t))$ at each point of G). Hence by Theorem 2, there exists a compact Hausdorff space G^* such that $B(G)$ is isometric and lattice isomorphic to the space $C(G^*)$ of all real valued continuous functions $x(t)$ defined on G^*. It is not difficult to see that in this case G^* is even a topological group, and that, in case G has sufficiently many almost periodic functions, G is a dense subgroup of G^*. Thus, for every topological group G with sufficiently many almost periodic functions, there exists a compact topological group G^* which contains G as a dense subgroup and such that every continuous almost periodic function defined on G can be uniquely extended to a continuous (and hence almost periodic) function defined on G^*. This is also a conse-

REMARK 2. There are two extreme cases of completely regular Hausdorff spaces, which seem to be interesting, namely, the case of compact Hausdorff spaces and the case of discrete spaces. These will be discussed in the following sections.

13. Case of compact Hausdorff spaces

In the arguments given above in section 12, consider the special case when the given Hausdorff space Ω' is compact. Then the homeomorphic image $\varphi(\Omega')$ of Ω' must also be compact, and consequently Ω must coincide with $\varphi(\Omega')$; i.e. Ω is homeomorphic to Ω'. This means that, if we start from a compact Hausdorff space Ω' and consider the abstract (M)-space $C(\Omega')$ of all bounded continuous real-valued functions defined on Ω', then the compact Hausdorff space Ω obtained in Theorem 2 is homeomorphic to the original space Ω', or in other words:

THEOREM 6. *The compact Hausdorff space Ω obtained in Theorem 2 is uniquely determined up to a homeomorphism.*

We can also state this theorem in the following form:

THEOREM 7. *Let Ω_1 and Ω_2 be two compact Hausdorff spaces. Then, in order that the two abstract (M)-spaces $C(\Omega_1)$ and $C(\Omega_2)$ be isometric and lattice-isomorphic, it is necessary and sufficient that Ω_1 and Ω_2 be homeomorphic to each other.*

Incidentally, we have also proved the following

THEOREM 8. *Let Ω be a compact Hausdorff space and let $C(\Omega)$ be a Banach space of all bounded continuous real-valued functions $x(t)$ defined on Ω. Then, every bounded linear functional $f_0(x)$ defined on $C(\Omega)$ which satisfies the conditions:*

$$\|f_0\| = 1, \tag{13.1}$$

$$x \geqq 0 \quad \textit{implies} \quad f_0(x) \geqq 0, \tag{13.2}$$

$$x \wedge y = 0 \quad \textit{implies} \quad f_0(x) = 0 \quad \textit{or} \quad f_0(y) = 0, \tag{13.3}$$

(i.e., *all conditions* (7.1)–(7.7)), *can be uniquely expressed in the form*:

$$f_0(x) = x(t_0), \tag{13.4}$$

where t_0 is a point of Ω.

Let us now consider a completely additive non-negative set function $\mu(E)$ defined for all Borel sets E of Ω such that $\mu(\Omega) = 1$. Then the integral

$$f_\mu(x) = \int_\Omega x(t)\mu(dt) \tag{13.5}$$

---

quence of a general duality theorem of Tannaka (Dualität der nichtkommutativen Gruppen, Tohoku Math. Journ., 45(1939), 1–12). The compactification of G with respect to a single almost periodic function (in this case we have to take a factor group G/N of G by a certain invariant subgroup N of G and then make this factor group G/N into a compact group $(G/N)^*$) was discussed by A. Weil (Sur les fonctions presque périodiques de von Neumann, C. R. Paris, 200(1935), 38–40) and the general case may be obtained from this by applying the idea of uniform space. (As for uniform spaces see A. Weil, Sur les espaces a structure uniforme et sur la topologie générale, Actualité, 551, Paris, 1937). It may be noticed that, after all, the method of uniform space and that of weak topology amount to the same thing.

will define a bounded linear functional $f_\mu(x)$ defined on $C(\Omega)$. It is easy to see that this $f_\mu(x)$ satisfies the conditions (13.1) and (13.2).

Moreover, we can prove that conversely every such functional is of the form (13.5); i.e.,

THEOREM 9. *Let Ω be a compact Hausdorff space, and let $C(\Omega)$ be defined as usual. Then every bounded linear functional $f(x)$ defined on $C(\Omega)$ which satisfies the conditions* (13.1) *and* (13.2) *can be expressed in the form* (13.5), *where $\mu(E)$ is a completely additive non-negative set funtion defined for all Borel sets E of Ω such that $\mu(\Omega) = 1$.*

This theorem was proved by S. Saks [15] under the assumption that Ω is a compact metric space.

PROOF.[11] For any open set O of Ω, put

$$\mu(O) = \text{l.u.b. } f(x), \tag{13.6}$$

where l.u.b. means the least upper bound for all continuous functions $x(t) \in C(\Omega)$ such that $x(t) \leqq 1$ for $t \in O$ and $x(t) = 0$ for $t \in \Omega - O$. It is clear that $\mu(\Omega) = 1$ and that $O \subseteq P$ implies $\mu(O) \leqq \mu(P)$ for any open sets O and P. It is also not difficult to see that we have $\mu(O + P) \leqq \mu(O) + \mu(P)$[12] for any open sets O and P, where the equality holds if O and P are disjoint. Next, for any subset M of Ω, put

$$\mu^*(M) = \text{g.l.b. } \mu(O), \tag{13.7}$$

where g.l.b. means the greatest lower bound for all open sets O which contain M. It is clear that for open sets O we have $\mu^*(O) = \mu(O)$. We shall first prove that $\mu^*(M)$ is a Carathéodory outer measure and that every open set is μ^*-measurable (in the sense of Carathéodory). Among the axioms of Carathéodory we have only to prove that

$$\mu^*\left(\textstyle\sum_{n=1}^\infty M_n\right) \leqq \textstyle\sum_{n=1}^\infty \mu^*(M_n) \tag{13.8}$$

for any sequence of subsets $\{M_n\}$ $(n = 1, 2, \cdots)$ of Ω such that $\sum_{n=1}^\infty \mu^*(M_n) < \infty$. It is easy to see that we have only to prove (13.8) for the special case when all M_n are open, i.e.,

$$\mu\left(\textstyle\sum_{n=1}^\infty O_n\right) \leqq \textstyle\sum_{n=1}^\infty \mu(O_n) \tag{13.9}$$

[11] The proof given below follows the idea of Professor von Neumann. See his lecture note on Haar measure [14].

[12] Let $\epsilon > 0$ be an arbitrary positive number. Then there exists an $x(t) \in C(\Omega)$ such that $f(x) > \mu(O + P) - \epsilon$, $x(t) \leqq 1$ for $t \in O + P$ and $x(t) = 0$ for $t \in \Omega - (O + P)$. Let C be the set of all $t \in \Omega$ at which $x(t) \geqq \epsilon$. Then C is a closed set contained in $O + P$. Hence C is a sum of two closed (not necessarily disjoint) sets C_1 and C_2 : $C = C_1 + C_2$ such that $C_1 \subseteq O$ and $C_2 \subseteq P$. Consequently, there exist two continuous functions $y(t)$, $z(t) \in C(\Omega)$ such that $y(t) = 1$ for $t \in C_1$, $0 \leqq y(t) \leqq 1$ for $t \in O - C_1$, $y(t) = 0$ for $t \in \Omega - O$; $z(t) = 1$ for $t \in C_2$, $0 \leqq z(t) \leqq 1$ for $t \in P - C_2$, $z(t) = 0$ for $t \in \Omega - P$. Then it is clear that $x(t) \leqq y(t) + z(t) + \epsilon$ for all $t \in \Omega$ and $f(y) \leqq \mu(O)$, $f(z) \leqq \mu(P)$. Consequently, $\mu(O + P) - \epsilon < f(x) \leqq f(y) + f(z) + \epsilon \leqq \mu(O) + \mu(P) + \epsilon$. Since $\epsilon > 0$ is arbitrary this proves the required inequality $\mu(O + P) \leqq \mu(O) + \mu(P)$.

for any sequence of open sets $\{O_n\}$ $(n = 1, 2, \cdots)$ such that $\sum_{n=1}^{\infty}\mu(O_n) < \infty$. Indeed, for any $\epsilon > 0$, let O_n be an open set such that $O_n \supseteq M_n$ and $\mu(O_n) < \mu^*(M_n) + 2^{-n}\cdot\epsilon$. Then we have $\sum_{n=1}^{\infty}O_n \supseteq \sum_{n=1}^{\infty}M_n$ and $\mu^*(\sum_{n=1}^{\infty}M_n) \leqq \mu(\sum_{n=1}^{\infty}O_n) \leqq \sum_{n=1}^{\infty}\mu(O_n) < \sum_{n=1}^{\infty}\mu^*(M_n) + \epsilon$. Since $\epsilon > 0$ is arbitrary we have (13.8).

Now, in order to prove (13.9), let $\epsilon > 0$ be an arbitrary positive number. Then there exists an $x(t) \in C(\Omega)$ such that $f(x) > \mu(\sum_{n=1}^{\infty}O_n) - \epsilon$, $x(t) \leqq 1$ for $t \in \sum_{n=1}^{\infty}O_n$ and $x(t) = 0$ for $t \in \Omega - \sum_{n=1}^{\infty}O_n$. If we denote by C the set of all points t at which $x(t) \geqq \epsilon$, then C is a closed set contained in $\sum_{n=1}^{\infty}O_n$. Hence there exists an integer N such that $C \subseteq \sum_{n=1}^{N}O_n$. Let now $y(t) \in C(\Omega)$ be such that $y(t) = 1$ for $t \in C$, $y(t) = 0$ for $t \in \Omega - \sum_{n=1}^{N}O_n$, and $0 \leqq y(t) \leqq 1$, elsewhere in Ω. Then we have $f(y) \leqq \mu(\sum_{n=1}^{N}O_n)$ and $x(t) \leqq y(t) + \epsilon$ for all $t \in \Omega$. Hence we have $\mu(\sum_{n=1}^{\infty}O_n) - \epsilon < f(x) \leqq f(y) + \epsilon \leqq \mu(\sum_{n=1}^{N}O_n) + \epsilon \leqq \sum_{n=1}^{N}\mu(O_n) + \epsilon \leqq \sum_{n=1}^{\infty}\mu(O_n) + \epsilon$. Since $\epsilon > 0$ is arbitrary we have (13.9).

In order to prove that every open set is μ^*-measurable, we have only to show that

$$\mu^*(M) \geqq \mu^*(M\cdot O) + \mu^*(M - M\cdot O) \tag{13.10}$$

for any open set O and for any subset $M \subseteq \Omega$. It is again easy to see that we have only to prove that (13.10) is true when M is open, i.e.

$$\mu(P) \geqq \mu(P\cdot O) + \mu^*(P - P\cdot O) \tag{13.11}$$

for any open sets O and P. Indeed, for any $\epsilon > 0$, there exists an open set $P \supseteq M$ such that $\mu(P) < \mu^*(M) + \epsilon$. Then we have $\mu^*(M) + \epsilon > \mu(P) \geqq \mu(P\cdot O) + \mu^*(P - P\cdot O) \geqq \mu^*(M\cdot O) + \mu^*(M - M\cdot O)$, and since $\epsilon > 0$ is arbitrary, we have (13.10).

To prove (13.11), let $\epsilon > 0$ be an arbitrary positive number. Then there exists an $x(t) \in C(\Omega)$ such that $f(x) > \mu(P\cdot O) - \epsilon$, $x(t) \leqq 1$ for $t \in P\cdot O$ and $x(t) = 0$ for $t \in \Omega - P\cdot O$. Let C be a set of all points t at which $x(t) \geqq \epsilon$. Then C is a closed set contained in $P\cdot O$. Let now $y(t) \in C(\Omega)$ be such that $f(y) > \mu(P - C) - \epsilon$, $y(t) \leqq 1$ for $t \in P - C$ and $y(t) = 0$ for $t \in \Omega - (P - C) = C + (\Omega - P)$. Then we have $x(t) + y(t) \leqq 1 + \epsilon$ for $t \in P$ and $x(t) + y(t) = 0$ for $t \in \Omega - P$. Hence we have $f(x) + f(y) \leqq \mu(P) + \epsilon$ and consequently $\mu^*(P - P\cdot O) + \mu(P\cdot O) - 2\epsilon \leqq \mu(P - C) + \mu(P\cdot O) - 2\epsilon < f(x) + f(y) < \mu(P) + \epsilon$. Since $\epsilon > 0$ is arbitrary we have (13.11).

Thus we have proved that μ^* is a Carathéodory outer measure and every open set is μ^*-measurable. Consequently every Borel set is μ^*-measurable, and if we put $\mu(E) = \mu^*(E)$ for Borel sets E, then $\mu(E)$ is the required measure. Since it is clear that $\mu(E)$ is completely additive and that $\mu(\Omega) = 1$, we have only to prove that we have

$$f(x) = \int_{\Omega} x(t)\mu(dt) \tag{13.12}$$

for any continuous function $x(t) \in C(\Omega)$.

Let $\epsilon > 0$ be an arbitrary positive number and let $\alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_n$ be a system of real numbers such that $\alpha_0 < - \| x \|$, $\alpha_n > \| x \|$ and $\alpha_i - \alpha_{i-1} < \epsilon$ for $i = 1, 2, \cdots, n$. Let E_i be the set of points t at which $x(t) > \alpha_i (i = 1, 2, \cdots, n)$. Then, each E_i is open, and if we put

$$x_i(t) \begin{cases} = 0, & \text{if } x(t) < \alpha_{i-1}, \\ = (\alpha_i - \alpha_{i-1})^{-1}(x(t) - \alpha_{i-1}), & \text{if } \alpha_{i-1} \leqq x(t) \leqq \alpha_i, \\ = 1 & \text{if } x(t) > \alpha_i, \end{cases}$$

then $x_i(t) \leqq 1$ for $t \in E_{i-1}$, $x_i(t) = 0$ for $t \in \Omega - E_{i-1}$ and $x(t) = \alpha_0 + \sum_{i=1}^n (\alpha_i - \alpha_{i-1})x_i(t)$. Hence we have (by (13.6)) $f(x_i) \leqq \mu(E_{i-1})$ and consequently (since $\mu(E_0) = 1$ and $\mu(E_n) = 0$)

$$\begin{aligned} f(x) &= \alpha_0 + \textstyle\sum_{i=1}^n(\alpha_i - \alpha_{i-1})f(x_i) \\ &\leqq \alpha_0 + \textstyle\sum_{i=1}^n(\alpha_i - \alpha_{i-1})\mu(E_{i-1}) \\ &= \textstyle\sum_{i=1}^n \alpha_i(\mu(E_{i-1}) - \mu(E_i)) \\ &\leqq \textstyle\sum_{i=1}^n \alpha_{i-1}(\mu(E_{i-1}) - \mu(E_i)) + \epsilon \\ &\leqq \int_\Omega x(t)\mu(dt) + \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we have $f(x) \leqq \int_\Omega x(t)\mu(dt)$. Since the inverse inequality follows by replacing $x(t)$ by $-x(t)$, the proof of Theorem 9 is completed.

REMARK 1. J. v. Neumann [14] started from a set function $\lambda(C)$ defined for compact sets C, and defined $\mu(O)$ for open sets O as follows:

(13.13) $$\mu(O) = \text{l.u.b. } \lambda(C),$$

where l.u.b. means the least upper bound for all compact sets C contained in O. In our case, we can, indeed, define $\lambda(C)$ by

(13.14) $$\lambda(C) = \text{g.l.b. } f(x),$$

where g.l.b. means the greatest lower bound for all $x(t) \in C(\Omega)$ such that $x(t) = 1$ for $t \in C$ and $x(t) \geqq 0$ for $t \in \Omega - C$. Then it is easy to see that the formula (13.13) gives the same value of $\mu(O)$ as (13.6).

REMARK 2. A completely additive non-negative set-function $\mu(E)$ is called *regular*, if there exist, for any Borel set E and for any $\epsilon > 0$ a closed set $C \subseteq E$ and an open set $O \supseteq E$ such that $\mu(C) > \mu(E) - \epsilon$, $\mu(O) < \mu(E) + \epsilon$. (If $\mu(E)$ is not non-negative, this last condition must be replaced by $| \mu(E) - \mu(E') | < \epsilon$ for any Borel set E' such that $C \subseteq E' \subseteq O$.)

In Theorem 9, the uniqueness of $\mu(E)$ was not discussed. If, however, we require that $\mu(E)$ be regular, then $\mu(E)$ is uniquely determined. Indeed, if there exist two completely additive regular set-functions $\mu_1(E)$ and $\mu_2(E)$ which satisfy (13.12) for any $x(t) \in C(\Omega)$, then $\mu(E) = \mu_1(E) - \mu_2(E)$ is also a completely additive regular set-function which satisfies

$$\int_\Omega x(t)\mu(dt) = 0$$

for any continuous function $x(t) \in C(\Omega)$. We shall show that we have $\mu(E) = 0$ for any Borel set E. Let E be an arbitrary Borel set. Then, for any $\epsilon > 0$, there exist a closed set $C \subseteq E$ and an open set $O \supseteq E$ such that $|\mu(E) - \mu(E')| < \epsilon$ for any Borel set E' with $C \subseteq E' \subseteq O$. Let now $x(t) \in C(\Omega)$ be a continuous function such that $x(t) = 1$ for $t \in C$, $x(t) = 0$ for $t \in \Omega - O$, and $0 \leqq x(t) \leqq 1$ elsewhere in Ω. Then, denoting by $x_E(t)$ the characteristic function of E, we have

$$\begin{aligned} |\mu(E)| &= \left| \mu(E) - \int_\Omega x(t)\mu(dt) \right| \\ &= \left| \int_\Omega (x_E(t) - x(t))\mu(dt) \right| \\ &\leqq \int_{O-C} |x_E(t) - x(t)|\,|\mu(dt)| \\ &\leqq \int_{O-C} |\mu(dt)| \leqq \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we must have $\mu(E) = 0$ for any Borel set E, and the uniqueness of $\mu(E)$ in Theorem 9 is proved.

Since, as is easily seen, every bounded linear functional defined on $C(\Omega)$ can be expressed as a difference of two positive bounded linear functionals, the result obtained in Theorem 9 can be extended to a following form:

THEOREM 10. *Let Ω be a compact Hausdorff space, and let $C(\Omega)$ be defined as usual. Let us further denote by $\mathfrak{M}(\Omega)$ the space of all completely additive regular set-functions $\mu(E)$ defined for all Borel sets E of Ω. If we put $\|\mu\|$ = total variation of $\mu(E) = \text{l.u.b.}_{E\subseteq\Omega}\,\mu(E) - \text{g.l.b.}_{E\subseteq\Omega}\,\mu(E)$, and $\mu \geqq \nu$ if and only if $\mu(E) \geqq \nu(E)$ for any Borel set $E \subseteq \Omega$, then $\mathfrak{M}(\Omega)$ is isometric and lattice isomorphic to the conjugate space $C^*(\Omega)$ of $C(\Omega)$. The correspondence between $C^*(\Omega)$ and $\mathfrak{M}(\Omega)$ is given by* (13.12).

We omit the proof.

EXAMPLE 1. Let Ω_0 be a closed interval: $0 \leqq t \leqq 1$. Then, as is well known, $C^*(\Omega_0)$ is a space of all functions of bounded variation (with total variation as its norm). It is easy to see that this space is nothing but the space $\mathfrak{M}(\Omega_0)$ defined in Theorem 10.

EXAMPLE 2. Let Ω be a compact Hausdorff space consisting of $t_1, t_2, \cdots, t_n, \cdots$ and t_∞, where each t_n ($n = 1, 2, \cdots$) is isolated and t_∞ is a limiting point of all others. Then $C(\Omega)$ is the space (C_0) of all convergent sequences of real numbers $x = (x_1, x_2, \cdots, x_n, \cdots, \lim_n x_n = x_\infty)$, with $\|x\| = \text{l.u.b.}_n |x_n|$ as its norm. In this case $C^*(\Omega)$ is the space of all sequences of real numbers $f = (f_1, f_2, \cdots, f_n, \cdots, f_\infty)$ such that $\|f\| = \sum_{n=1}^\infty |f_n| + |f_\infty| < \infty$.

In Theorem 9 we have seen that every positive bounded linear functional defined on $C(\Omega)$ can be considered as a measure $\mu(E)$ defined for all Borel sets E of Ω. And, as is easily seen, Theorem 8 means that if we assume the additional

condition (13.3) then this measure $\mu(E)$ is concentrated in one point t_0 of Ω. Moreover, we can show the following:

THEOREM 11. *Let Ω be a compact Hausdorff space and let $C(\Omega)$ be a space of all bounded continuous real-valued functions $x(t)$ defined on Ω. Then every bounded linear functional $f_\mu(x)$ of the form (13.5) can be approximated (in the sense of weak topology) by convex combinations $f_\Sigma(x)$ of the functionals of the form (13.4):*

$$f_\Sigma(x) = \sum_{p=1}^{k} \alpha_p x(t_p), \tag{13.15}$$

where $t_p \in \Omega$, $\alpha_p \geqq 0$ $(p = 1, 2, \cdots, k)$ and $\sum_{p=1}^{k} \alpha_p = 1$.

In other words, if we denote by Δ_0 and Δ_μ the set of all bounded linear functionals $f_0(x)$ and $f_\mu(x)$ of the form (13.4) and (13.5) respectively, then Δ_μ is contained in (hence coincides with) the convex closure of Δ_0 with respect to the weak topology.

PROOF. Let $f_\mu(x)$ be a bounded linear functional which is defined by (13.5), where $\mu(E)$ is a completely additive non-negative set-function defined for all Borel sets E of Ω with $\mu(\Omega) = 1$. Then, for any $x(t) \in C(\Omega)$ and for any $\epsilon > 0$, there exists a decomposition $\Omega = \sum_{p=1}^{k} E_p$, $E_p \cdot E_q = 0$ $(p \neq q)$ of Ω into disjoint Borel sets such that

$$| f_\mu(x) - \sum_{p=1}^{k} x(t_p)\mu(E_p) | < \epsilon$$

for any $t_p \in E_p$, $p = 1, 2, \cdots, k$. Let now $x_i(t) \in C(\Omega)$ $(i = 1, 2, \cdots, n)$ be an arbitrary finite system and let $\Omega = \sum_{p=1}^{k_i} E_p^{(i)}$, $E_p^{(i)} \cdot E_q^{(i)} = 0$ $(p \neq q)$ be the corresponding decomposition of Ω into Borel sets (with the same ϵ), then the decomposition: $\Omega = \sum_{p_1=1}^{k_1} \sum_{p_2=1}^{k_2} \cdots \sum_{p_n=1}^{k_n} E_{p_1 p_2 \ldots p_n}$, where $E_{p_1 p_2 \cdots p_n} = E_{p_1}^{(1)} E_{p_2}^{(2)} \cdots E_{p_n}^{(n)}$, will satisfy

$$|f_\mu(x_i) - \sum_{p_1=1}^{k_1} \sum_{p_2=1}^{k_2} \cdots \sum_{p_n=1}^{k_n} x(t_{p_1 p_2 \ldots p_n}) \mu(E_{p_1 p_2 \ldots p_n}) | < \epsilon$$

for any $t_{p_1 p_2 \cdots p_n} \in E_{p_1 p_2 \cdots p_n}$ $(1 \leqq p_i \leqq k_i, i = 1, 2, \cdots, n)$ and for any x_i $(i = 1, 2, \cdots, n)$. If we now put $\alpha_{p_1 p_2 \cdots p_n} = \mu(E_{p_1 p_2 \cdots p_n})$, then this fact means that the bounded linear functional

$$f_\Sigma(x) = \sum_{p_1=1}^{k_1} \sum_{p_2=1}^{k_2} \cdots \sum_{p_n=1}^{k_n} \alpha_{p_1 p_2 \cdots p_n} f(t_{p_1 p_2 \cdots p_n}) \tag{13.16}$$

lies inside a weak neighborhood $U(f_\mu; x_1, x_2, \cdots, x_n; \epsilon)$ of f_μ. Consequently, f_μ is a limiting point of all functionals $f_\Sigma(x)$ of the form (13.7). Since $1 = \sum_{p_1=1}^{k_1} \sum_{p_2=1}^{k_2} \cdots \sum_{p_n=1}^{k_n} \alpha_{p_1 p_2 \cdots p_n}$, this proves Theorem 11.

REMARK. This theorem is also true even if Ω is not compact.

14. Case of discrete spaces, 1. Theorem of M. H. Stone

In this section we shall consider the case when Ω' is discrete, i.e., when every point of Ω' is an isolated point. Any two such discrete spaces with the same power are clearly homeomorphic to each other. Consequently, the compact Hausdorff space Ω which is obtained in Theorem 5, is uniquely determined (up to a homeomorphism) by the power $\mathfrak{m}$ of the discrete space Ω. We shall denote this space by $\Omega(\mathfrak{m})$.

Since the space Ω' is discrete by assumption, every bounded real-valued function is continuous on Ω'. Let now $x_E(t)$ be a characteristic function of an arbitrary subset E of Ω'. Then $x_E(t)$ is continuous on Ω' and consequently $x_E(t)$ can be uniquely extended to a continuous function $x_E(f)$ defined on Ω. It is clear that this function $x_E(f)$ takes only the values 0 and 1. Indeed, since $\varphi(\Omega')$ is dense in Ω, there exists, for any $\epsilon > 0$, a point $t \in \Omega'$ such that $| x_E(f) - x_E(t) | < \epsilon$. Consequently, since $x_E(t)$ takes only the values 0 and 1, we must have $x_E(f) = 0$ or $= 1$ for any $f \in \Omega$.

Thus we have seen that $x_E(f)$ is also a characteristic function of a certain subset of Ω. Denote this set by $\bar{E}$. It is clear that $\bar{E}$ is open-and-closed, and if we consider Ω' as a subset of Ω (this is possible since Ω' is homeomorphic to $\varphi(\Omega') \subseteq \Omega$), then we have $E \subseteq \bar{E}$. Moreover, it is easy to see that in case E is reduced to one point, $\bar{E}$ must also be reduced to the same point.

Conversely, let $\bar{E}$ be an arbitrary open-and-closed subset of Ω. Then the characteristic function $\bar{x}_{\bar{E}}(f)$ of $\bar{E}$ is continuous on Ω, and the part $x_E(t)$ of $\bar{x}(_{\bar{E}}f)$ on Ω' must take only the values 0 and 1; i.e., $x_E(t)$ is also a characteristic function of a certain subset of Ω'. Denote this set by E. Then it is easy to see that the correspondence $E \leftrightarrow \bar{E}$ is one-to-one. Moreover, as is easily seen, this mapping $E \leftrightarrow \bar{E}$ is isomorphic; i.e., $E \leftrightarrow \bar{E}$, $F \leftrightarrow \bar{F}$ imply $E + F \leftrightarrow \bar{E} + \bar{F}$, $E \cdot F \leftrightarrow \bar{E} \cdot \bar{F}$ and $\Omega' - E \leftrightarrow \Omega - \bar{E}$.

This result is nothing but the theorem of M. H. Stone [17] on the concrete representation of a Boolean algebra $\mathfrak{B}$ for the special case when $\mathfrak{B}$ is the Boolean algebra of all subsets of a space Ω'. But we can proceed further, as Stone did, to the general case of an abstract Boolean algebra with a unit element. (The case of a Boolean algebra without a unit element can be treated by using Theorem 1. But, in order to avoid unnecessary confusions, we shall discuss only those cases when a unit element exists.)

In the following lines we shall give a general outline of our proof. Let $\mathfrak{B}$ be an abstract Boolean algebra with a unit element e, and let us consider a space $M(\mathfrak{B})$ of all step functions defined on $\mathfrak{B}$. Under a step function defined on $\mathfrak{B}$ we understand a system $\begin{pmatrix} a_1, & a_2, & \cdots, & a_n \\ \alpha_1, & \alpha_2, & \cdots, & \alpha_n \end{pmatrix}$ where $a_i \in \mathfrak{B}$ $(i = 1, 2, \cdots, n)$, $\sum_{i=1}^n a_i = e$, $a_i a_j = 0$ $(i \neq j)$ and $\alpha_i (i = 1, 2, \cdots, n)$ is a real number. Two step functions $\begin{pmatrix} a_1, & a_2, & \cdots, & a_m \\ \alpha_1, & \alpha_2, & \cdots, & \alpha_m \end{pmatrix}$ and $\begin{pmatrix} b_1, & b_2, & \cdots, & b_n \\ \beta_1, & \beta_2, & \cdots, & \beta_n \end{pmatrix}$ are considered as equal if and only if $\sum_{\alpha_i = \gamma} a_i = \sum_{\beta_j = \gamma} b_j$ for any real number γ, where $\sum_{\alpha_i = \gamma} a_i$, for example, means the sum of all a_i such that $\alpha_i = \gamma$. If we now put

$$\begin{pmatrix} a_1, & a_2, & \cdots, & a_m \\ \alpha_1, & \alpha_2, & \cdots, & \alpha_m \end{pmatrix} + \begin{pmatrix} b_1, & b_2, & \cdots, & b_n \\ \beta_1, & \beta_2, & \cdots, & \beta_n \end{pmatrix} = \begin{pmatrix} a_1 b_1, & a_1 b_2, & \cdots, & a_m b_n \\ \alpha_1 + \beta_1, & \alpha_1 + \beta_2, & \cdots, & \alpha_m + \beta_n \end{pmatrix}, \tag{14.1}$$

$$\lambda \begin{pmatrix} a_1, & a_2, & \cdots, & a_n \\ \alpha_1, & \alpha_2, & \cdots, & \alpha_n \end{pmatrix} = \begin{pmatrix} a_1, & a_2, & \cdots, & a_n \\ \lambda\alpha_1, & \lambda\alpha_2, & \cdots, & \lambda\alpha_n \end{pmatrix}, \tag{14.2}$$

$$\left\| \begin{pmatrix} a_1, a_2, \cdots, a_n \\ \alpha_1, \alpha_2, \cdots, \alpha_n \end{pmatrix} \right\| = \max_{1 \leq i \leq n} | \alpha_i \tag{14.3}$$

$$\begin{pmatrix} a_1, a_2, \cdots, a_n \\ \alpha_1, \alpha_2, \cdots, \alpha_n \end{pmatrix} \geqq 0 \quad \text{if and only if } \alpha_i \geqq 0 \text{ for } i = 1, 2, \cdots, n, \tag{14.4}$$

then $M(\mathfrak{B})$ satisfies all the axioms of abstract (M)-spaces except the completeness with respect to the norm (14.3). If we now make it complete with respect to this norm, then the space $\overline{M(\mathfrak{B})}$ thus obtained is surely an abstract (M)-space with $\begin{pmatrix} e \\ 1 \end{pmatrix}$ as its unit element. (Of course, we must define the partial ordering in $\overline{M(\mathfrak{B})}$ in a suitable manner). Moreover, if we put $x_a = \begin{pmatrix} a, & e - a \\ 1, & 0 \end{pmatrix}$, then $a \leftrightarrow x_a$ is a one-to-one mapping which embeds $\mathfrak{B}$ isomorphically into $\overline{M(\mathfrak{B})}$.

If we now apply Theorem 2 to this space $\overline{M(\mathfrak{B})}$, then we have a concrete representation of $\overline{M(\mathfrak{B})}$ as the space $C(\Omega)$ of all real valued continuous functions $x(f)$ defined on a compact Hausdorff space Ω. The unit element $\mathbf{1} = \begin{pmatrix} e \\ 1 \end{pmatrix}$ of $\overline{M(\mathfrak{B})}$ is represented by a constant function $x(f) \equiv 1$, and it is easy to see that, for any $a \in \mathfrak{B}$ the function $x_a(f) \in C(\Omega)$ which corresponds to $x_a \in \overline{M(\mathfrak{B})}$ is a characteristic function of a certain subset of Ω. This follows from the fact that $2x_a \wedge \mathbf{1} = x_a$ for any $a \in \mathfrak{B}$. Let us denote this set by E_a. E_a is clearly an open-and-closed subset of Ω and the correspondence $a \leftrightarrow E_a$ is an isomorphism. Thus $\mathfrak{B}$ is concretely represented as a Boolean algebra of open-and-closed subsets of Ω. Moreover, it is easy to see that Ω is a totally disconnected space, and that for any open-and-closed subset E of Ω, there exists an $a \in \mathfrak{B}$ such that $E_a = E$.

Thus we have proved Stone's

THEOREM 12. *For any abstract Boolean algebra $\mathfrak{B}$ with a unit element, there exists a totally disconnected compact Hausdorff space Ω such that $\mathfrak{B}$ is isomorphic to the Boolean algebra of all open-and-closed subsets of Ω.*

If we, however, wish to obtain a direct proof of this theorem, then we can proceed in the following way: let $\mathfrak{B}$ be an arbitrary Boolean algebra with unit e and consider the totality Ω^* of all real valued functions $f(x)$ defined on $\mathfrak{B}$ whose values are either 0 or 1. Ω^* is a compact Hausdorff space with respect to the weak topology.[13] Let us further consider a real valued function $f(x) \in \Omega^*$ which satisfies the following conditions:

$$f(0) = 0, \qquad f(e) = 1, \tag{14.5}$$

$$f(x \vee y) = \max (f(x), f(y)), \tag{14.6}$$

$$f(x \wedge y) = \min (f(x), f(y)), \tag{14.7}$$

$$f(x + y) = f(x) + f(y), \qquad \text{mod } 2, \tag{14.8}$$

$$f(xy) = f(x)f(y), \tag{14.9}$$

where $x \vee y = x + y + xy$, $x \wedge y = xy$. (These conditions are clearly not mutually independent.) These conditions mean that $x \to f(x)$ is a homomorphism of $\mathfrak{B}$ onto the Boolean algebra $\mathfrak{B}_0$ consisting of 0 and 1 ($\mathfrak{B}_0$ is the simplest non-trivial Boolean algebra). The totality Ω of all such $f(x)$ constitutes again a compact Hausdorff space with respect to the weak topology,[13] and if we put $x(f) = f(x)$ for any fixed $x \in \mathfrak{B}$, then $x(f)$ is a continuous function defined on Ω whose values are always 0 or 1. It is easy to see that $x \to x(f)$ is a homomorphic mapping. In order to show that this is an isomorphism, we have only to prove that for any $a \in \mathfrak{B}$, $b \in \mathfrak{B}$, $a \neq b$, there exists an $f_0 \in \Omega$ such that $f_0(a) \neq f_0(b)$. This fact corresponds to Lemma 7.1 in the discussions of Part II, and also to the existence of a prime ideal, which contains one of a and b but not the other, in the proof of M. H. Stone.

In order to prove the existence of such a function, let Σ be an arbitrary finite subset of $\mathfrak{B}$ which contains 0, 1, a and b. Then there exists a function $f_\Sigma(x) \in \Omega^*$ which satisfies (14.5)–(14.9) for any $x, y \in \Sigma$ and such that $f_\Sigma(a) \neq f_\Sigma(b)$. This follows from the fact that the Boolean sub-algebra $\mathfrak{B}_\Sigma$ of $\mathfrak{B}$ which is generated by Σ has also only a finite number of elements, and that for such a finite Boolean algebra $\mathfrak{B}_\Sigma$ Theorem 12 is trivially true. Let Ω_Σ be the set of all such $f_\Sigma(x)$. Then Ω_Σ is clearly a non-empty closed (and consequently compact) subset of Ω^*. Moreover, the family $\{\Omega_\Sigma\}$, where Σ runs all finite subsets of $\mathfrak{B}_1$ has a finite intersection property. This follows from the fact that $\prod_{i=1}^{n} \Omega_{\Sigma_i} \supseteq \Omega_{\Sigma_1+\Sigma_2+\cdots+\Sigma_n} \neq 0$ for any finite number of finite subsets Σ_i $(i = 1, 2, \cdots, n)$ of $\mathfrak{B}$. Hence there exists an element f_0 which belongs to all such Ω_Σ and this element f_0 is clearly the required one.

Thus we have proved that the Boolean algebra $\mathfrak{B}$ is isomorphically represented by the system of continuous functions $x(f) = f(x)$ (whose values are 0 or 1) defined on a compact space Ω. Let now E_a be the set of all $f(x) \in \Omega$ such that $f(a) = 1$. Then E_a is clearly closed, and since $\Omega - E_a = E_{e-a}$, E_a must also be open. It is easy to see that $a \leftrightarrow E_a$ is an isomorphism, and all that we have to do is to prove that Ω is totally disconnected and that for any open-and-closed subset E of Ω there exists an element $a \in \mathfrak{B}$ such that $E = E_a$. (This part of the proof corresponds to Lemma 7.2.)

To prove these, let us first show that for any $f \in \Omega$, $g \in \Omega$, $f \neq g$, there exists an open-and-closed subset E of Ω such that $f \in E$, $g \bar{\in} E$. Since $f \neq g$, there exists an $a_0 \in \mathfrak{B}$ such that $f(a_0) \neq g(a_0)$. Without loss of generality we may assume $f(a_0) = 1$, $g(a_0) = 0$, since otherwise we may replace a_0 by $e - a_0$. Then the open-and-closed subset E_{a_0} has the required property.

Next we shall show that for any $f \in \Omega$ and for any open set $U(f)$ which contains f, there exists an $a_0 \in \mathfrak{B}$ such that $f \in E_{a_0} \subseteq U(f)$. By the above result, for any $g \bar{\in} U(f)$, there exists an $a \in \mathfrak{B}$ such that $f \in E_a$, $g \bar{\in} E_a$. Since $\Omega - E_a$ is an open set which contains g, there exists by the compactness of $\Omega - U(f)$,

[13] For any $f_0 \in \Omega^*$ its weak neighborhood $U(f_0; x_1, x_2, \cdots, x_n)$ is defined as the totality of all $f \in \Omega^*$ such that $f(x_i) = f_0(x_i)$ for $i = 1, 2, \cdots, n$, where $\{x_i\}$ $(i = 1, 2, \cdots, n)$ is an arbitrary finite system of elements of $\mathfrak{B}$.

a finite number of elements $a_i \in \mathfrak{B}$ $(i = 1, 2, \cdots, n)$ such that $f \in E_{a_i}$ $(i = 1, 2, \cdots, n)$ and $\Omega - U(f) \subseteq \sum_{i=1}^{n} (\Omega - E_{a_i})$ i.e., $\prod_{i=1}^{n} E_{a_i} \subseteq U(f)$. If we now put $a_0 = a_1 \wedge a_2 \wedge \cdots \wedge a_n$, then we have $f \in E_{a_0} = \prod_{i=1}^{n} E_{a_i} \subseteq U(f)$. This proves the total disconnectedness of Ω.

Lastly, we shall prove that for any open-and-closed subset E of Ω, there exists an element $a_0 \in \mathfrak{B}$ such that $E = E_{a_0}$. For any $f \in E$, there exists, by the above result, an $a \in \mathfrak{B}$ such that $f \in E_a \subseteq E$. Since each E_a is open and since E is compact, there exists a finite number of elements $a_i \in \mathfrak{B}$ $(i = 1, 2, \cdots, n)$ such that $E = \sum_{i=1}^{n} E_{a_i}$. If we put $a_0 = a_1 \vee a_2 \vee \cdots \vee a_n$, then we have $E_{a_0} = \sum_{i=1}^{n} E_{a_i} = E$.

The proof of Theorem 12 is hereby completed.

15. Case of discrete spaces 2. Banach limits

Let (m) be a space of all bounded sequences of real numbers $x = \{x_n\}$, $n = 1, 2, \cdots$ with $\|x\| = \sup_n |x_n|$ as its norm. A functional $\operatorname{Lim} x_n$ defined on (m) is called a Banach limit, if it satisfies the following conditions.

$$\operatorname{Lim} (x_n + y_n) = \operatorname{Lim} x_n + \operatorname{Lim} y_n, \tag{15.1}$$

$$\operatorname{Lim} (\lambda x_n) = \lambda \operatorname{Lim} x_n, \tag{15.2}$$

$$x_n \geqq 0, \quad n = 1, 2, \cdots \quad \textit{imply} \quad \operatorname{Lim} x_n \geqq 0, \tag{15.3}$$

$$x_n = 1, \quad n = 1, 2, \cdots \quad \textit{imply} \quad \operatorname{Lim} x_n = 1, \tag{15.4}$$

$$\operatorname{Lim} x_{n+1} = \operatorname{Lim} x_n. \tag{15.5}$$

If we consider a discrete space Ω' consisting of a denumerable number of isolated points $\{t_n\}$ $(n = 1, 2, \cdots)$, then (m) is nothing but a space $C(\Omega')$ of all bounded continuous real-valued functions $x(t)$ defined on Ω' $(x_n = x(t_n), n = 1, 2, \cdots)$. If we now consider $C(\Omega')$ as an abstract (M)-space in the usual manner, then $C(\Omega')$ is isometric and lattice isomorphic to the space $C(\Omega)$, where $\Omega = \Omega(\aleph_0)$ is a compact space which was defined in the beginning of section 14. If we put $f(x) = \operatorname{Lim} x_n$, then the functional $f(x)$ defined on $C(\Omega) = C(\Omega')$ satisfies the following conditions:

$$f(x + y) = f(x) + f(y), \tag{15.6}$$

$$f(\lambda x) = \lambda f(x), \tag{15.7}$$

$$x \geqq 0 \quad \textit{implies} \quad f(x) \geqq 0, \tag{15.8}$$

$$f(\mathbf{1}) = 1, \quad \textit{where} \quad \mathbf{1} = (1, 1, \cdots), \tag{15.9}$$

$$f(T(x)) = f(x), \quad \textit{where} \quad T(x) = (x_2, x_3, \cdots) \text{ if } x = (x_1, x_2, \cdots). \tag{15.10}$$

Let us observe only the conditions (15.6)–(15.9) for the time being. Then these conditions mean that $f(x)$ is a positive bounded linear functional with $f(\mathbf{1}) = 1$. But this last condition is equivalent to $\|f\| = 1$ (see Lemma 11.1). Consequently, by Theorem 9, there exists a completely additive, regular non-

negative set function $\mu(E)$ defined for all Borel sets E of the space Ω with $\mu(\Omega) = 1$ such that

$$f(x) = \int_\Omega x(t)\mu(dt) \tag{15.11}$$

for any $x(t) \in C(\Omega) = C(\Omega')$. Moreover, if we require an additional condition:

$$x \wedge y = 0 \quad \textit{implies} \quad f(x) = 0 \quad \text{or} \quad f(y) = 0, \tag{15.12}$$

then the measure $\mu(E)$ is concentrated in one point $t_0 \in \Omega$. If $t_0 \in \Omega - \Omega'$, then the functional $f(x) = x(t_0)$ is not trivial. It is easy to see that in this case we have

$$f(x) = 0 \textit{ for any } x \textit{ of the form } x = (x_1, x_2, \cdots, x_n, 0, 0, \cdots). \tag{15.13}$$

REMARK. It will be easily seen that under the conditions (15.6)–(15.9) the condition (15.12) is equivalent to

$$(f(x))^2 = f(x^2), \tag{15.14}$$

where $x^2 = (x_1^2, x_2^2, \cdots, x_n^2, \cdots)$ if $x = (x_1, x_2, \cdots, x_n, \cdots)$. Such functionals $f(x)$ were discussed by J. von Neumann [14], and it is to be remarked that (15.14) is equivalent to

$$f(x)f(y) = f(x \cdot y), \tag{15.15}$$

where $xy = (x_1y_1, x_2y_2, \cdots, x_ny_n, \cdots)$ if $x = (x_1, x_2, \cdots, x_n, \cdots)$ and $y = (y_1, y_2, \cdots, y_n, \cdots)$. This last condition together with (15.6), (15.7) means that $x \to f(x)$ is a homomorphic mapping of a ring (m) onto a ring of real numbers.

Next we shall discuss the condition (15.10). Let us denote by Δ_μ the set of all functionals $f_\mu(x)$ of the form (15.11) with $\mu(E) \geqq 0$, $\mu(\Omega) = 1$, i.e., the functionals which satisfy (15.6)–(15.9). Then Δ_μ is a compact convex set with respect to the weak topology. If we put $\bar{f}(x) = f(T(x))$, then $f \to \bar{f} = \varphi(f)$ is an *affine* continuous mapping of Δ_μ into itself, where under affine mapping we mean a mapping φ such that $\varphi(\lambda f + \lambda' g) = \lambda\varphi(f) + \lambda'\varphi(g)$ for any $f, g \in \Delta_\mu$; $\lambda, \lambda' \geqq 0$, $\lambda + \lambda' = 1$. It is easy to see that the condition (15.10) means that f is a fixed point under this affine mapping φ of Δ_μ into itself, and the existence of such a fixed point follows from the following

LEMMA 15.1. *Let E be a Banach space, and let Δ be a convex compact set in the conjugate space E^* of E (with respect to the weak topology). If $\varphi(f)$ is a continuous affine mapping of Δ into itself, then there exists a fixed element $f_0 \in \Delta$ such that $\varphi(f_0) = f_0$.*

PROOF. Let Δ_n be the set of all elements of the form: $n^{-1}(f + \varphi(f) + \cdots + \varphi^{n-1}(f))$, $f \in \Delta$. Then Δ_n is a non-empty compact subset of Δ and the family $\{\Delta_n\}$ $(n = 1, 2, \cdots)$ has clearly a finite intersection property: for any n_i $(i = 1, 2, \cdots, k)$ $\Delta_{n_1} \cdot \Delta_{n_2} \cdot \cdots \cdot \Delta_{n_k} \neq 0$. This follows directly from $\Delta_{n_1} \cdot \Delta_{n_2} \cdot \cdots \cdot \Delta_{n_k} \supseteq \Delta_{n_1 n_2 \cdots n_k} \neq 0$. Hence, there exists an $f_0(x)$ which belongs to all Δ_n,

$n = 1, 2, \cdots$. It is easy to see that this f_0 is a required one. (For generalizations of this lemma, see [8].)

REMARK. We can also prove the existence of a fixed point in the folllowing way: Let $S(x)$ be a mapping defined on (m) by $x \to S(x) = x'$: $x = (x_1, x_2, \cdots, x_n, \cdots)$, $x' = (x_1, 2^{-1}(x_1 + x_2), \cdots, n^{-1}(x_1 + x_2 + \cdots + x_n), \cdots)$, and put $f_0(x) = f(S(x))$, where $f(x)$ is an arbitrary bounded linear functional which satisfies (15.6)–(15.9) and (15.13). Then we have $f_0(T(x)) = f_0(x)$. Indeed, writing $f(x) = f(x_1, x_2, \cdots, x_n, \cdots)$ for $x = (x_1, x_2, \cdots) \in (m)$, we have $f_0(x) - f_0(T(x)) = f(S(x)) - f(S(T(x))) = f(x_1, 2^{-1}(x_1 + x_2), \cdots, n^{-1}(x_1 + x_2 + \cdots + x_n) \cdots) - f(x_2, 2^{-1}(x_2 + x_3), \cdots, n^{-1}(x_2 + x_3 + \cdots + x_{n+1}), \cdots) = f(0, 0, \cdots, 0, (n + 1)^{-1}(x_1 + x_2 + \cdots + x_{n+1}), (n + 2)^{-1}(x_1 + x_2 + \cdots + x_{n+2}), \cdots) - f(0, 0, \cdots, 0, (n + 1)^{-1}(x_2 + x_3 + \cdots + x_{n+2}), (n + 2)^{-1}(x_2 + x_3 + \cdots + x_{n+3}), \cdots) = f(0, 0, \cdots, 0, (n + 1)^{-1}(x_1 - x_{n+2}), (n + 2)^{-1}(x_1 - x_{n+3}), \cdots)$. Hence, $| f_0(x) - f_0(T(x)) | \leqq f(0, 0, \cdots, 0, 2(n + 1)^{-1}\| x \|, 2(n + 2)^{-1}\| x \|, \cdots) \leqq 2n^{-1} \| x \|$ for $n = 1, 2, \cdots$. Since n is arbitrary, we must have $f_0(x) = f_0(T(x))$.

Summing up the results, we have

THEOREM 13. *Every Banach limit which satisfies* (15.1)–(15.5) *can be considered as a bounded linear functional $f(x)$ defined on a Banach space $C(\Omega(\aleph_0))$ which satisfies the conditions* (15.6)–(15.10). (*For the definition of $\Omega(\aleph_0)$ see the beginning of section* 14.) *The functionals $f(x)$ which satisfy* (15.6)–(15.9) *can be expressed as an integral* (15.11). *The totality of all such functionals forms a compact convex set Δ_μ in the conjugate space $C^*(\Omega(\aleph_0))$ of $C(\Omega(\aleph_0))$, and every Banach limit* (*which satisfies the additional condition* (15.10)) *can be obtained as a fixed point of the affine mapping φ of Δ_μ into itself.*

It is easy to see that the result of this section can be extended to the case where the power of a given discrete space Ω' is greater than $\aleph_0$.

16. Relations with abstract (L)-spaces

A Banach space (AL) is called an *abstract (L)-space* if it is a linear lattice, and if it satisfies, besides the axioms (1.1)–(1.8), the following condition:

$$(16.1) \qquad x \geqq 0, \quad y \geqq 0 \quad \textit{imply} \quad \| x + y \| = \| x \| + \| y \|.$$

It is easy to see that the condition (1.16) is also satisfied. Such a Banach space was introduced by G. Birkhoff [1], and it was shown that a large number of results concerning dependent probabilities can be extended to such spaces.

THEOREM 14. *For any abstract (L)-space (AL) with an F-unit* (*see section* 3), *there exists a totally disconnected compact Hausdorff space Ω and a completely additive measure $m(E)$ defined for all Borel sets E of Ω such that (AL) is isometric and lattice isomorphic to the space $L(\Omega; m)$ of all measurable functions $x(t)$ which are integrable with respect to $m(E)$ on Ω* $\left(\| x \| = \int_\Omega | x(t) | \, m(dt);\ x \geqq y\colon x(t) \geqq y(t) \textit{ almost everywhere on } \Omega\right)$.

This theorem was proved in [9], [11].

REMARK 1. In my former paper [11] it was shown that $m(E)$ is defined on the smallest Borel field which contains all open-and-closed subsets of Ω. It is, however, easy to see that this measure can be extended to a completely additive measure $m'(E)$ defined for all Borel sets E of Ω. Moreover, this extension is unique if we require that $m'(E)$ is regular, and it is easy to see that we have $L(\Omega; m) = L(\Omega; m')$. This follows from the fact that, for any Borel set E and for any $\epsilon > 0$, there exists an open-and-closed set E' such that $m'(E + E' - E \cdot E') < \epsilon$. (See Remark 2, after Theorem 9.)

REMARK 2. Theorem 14 is true in the following sense even if there exists no F-unit in (AL). In this case, (AL) is decomposed into a direct sum of abstract (L)-spaces $(AL)_\alpha$ with an F-unit ($\alpha \in \mathfrak{M}$, where $\mathfrak{M}$ is a non-denumerable set of indices α), and each $(AL)_\alpha$ is concretely represented as in Theorem 14. Hence there exists a family of compact Hausdorff spaces Ω_α, $\alpha \in \mathfrak{M}$ such that $(AL) = \sum_{\alpha \in \mathfrak{M}} (AL)_\alpha$ is isometric and lattice isomorphic to $L(\Omega; m) = \sum_{\alpha \in \mathfrak{M}} L(\Omega_\alpha; m)$, where $\Omega = \sum_{\alpha \in \mathfrak{M}} \Omega_\alpha$ and each element of $L(\Omega; m)$ is a measurable function $x(t)$ defined on $\Omega = \sum_{\alpha \in \mathfrak{M}} \Omega_\alpha$ such that there exists a sequence of indices α_n, $n = 1, 2, \cdots$ such that $\| x \| = \sum_{n=1}^{\infty} \int_{\Omega \alpha_n} | x(t) | m(dt) < \infty$ and $x(t) = 0$ almost everywhere on Ω_α, $\alpha \neq \alpha_n$, $n = 1, 2, \cdots$. (For details, see [11].)

We shall discuss in this chapter the relations between abstract (M)-spaces and abstract (L)-spaces. We begin with

LEMMA 16.1. *Let E be a Banach lattice, i.e., a Banach space with a partial ordering relation $x \geqq y$ which satisfies the conditions* (1.1)–(1.8) *and* (1.16). *Then the conjugate space E^* of E is also a Banach lattice, if we put $f \geqq g$ if and only if $f(x) \geqq g(x)$ for any $x \geqq 0$, $x \in E$.*

PROOF. We have only to show that the conditions (1.5), (1.8) and (1.16) are satisfied for the partial ordering $f \geqq g$. To define $h = f \vee g$, put $h(x) =$ l.u.b. $\{f(x_1) + g(x_2)\}$ for $x \geqq 0$, where l.u.b. means the least upper bound for all decompositions $x = x_1 + x_2$, $x_1 \geqq 0$, $x_2 \geqq 0$. $h(x)$ is thus defined for $x \geqq 0$, and if we put $h(x) = h(x \vee 0) - h((-x) \vee 0)$, then this $h(x)$ is a required one (see, for example, H. Freudenthal [5]).

Next, to show that $f \geqq g \geqq 0$ implies $\|f\| \geqq \|g\| \geqq 0$, let $\epsilon > 0$ be an arbitrary positive number. Then there exists an $x \in E$, such that $\| x \| \leqq 1$ and $g(x) > \| g \| - \epsilon$. We may assume that $x \geqq 0$, since otherwise we can replace x by $x \vee 0$. Indeed, we have $\| x \vee 0 \| \leqq \| x \vee 0 + (-x) \vee 0 \| = \| x \| \leqq 1$ and $g(x \vee 0) \geqq g(x) > \| g \| - \epsilon$. Consequently, we have $f(x) \geqq g(x) > \| g \| - \epsilon$ for an x with $\| x \| \leqq 1$, and this implies $\|f\| \geqq \|g\| - \epsilon$. Since $\epsilon > 0$ is arbitrary, we have $\|f\| \geqq \|g\|$.

Finally, to prove that $f \wedge g = 0$ implies $\|f + g\| = \|f - g\|$, let $\epsilon > 0$ be an arbitrary positive number. Then there exists an $x \in E$ such that $\| x \| \leqq 1$ and $f(x) + g(x) > \|f + g\| - \epsilon$. We may again assume that $x \geqq 0$. Since $f \wedge g = 0$, there exists a decomposition $x = x_1 + x_2$, $x_1 \geqq 0$, $x_2 \geqq 0$ such that

$0 \leqq f(x_1) + g(x_2) < \epsilon$.[14] If we now put $x' = -x_1 + x_2$, then we have $\| x' \| = \| (x_1 - x_2) \vee 0 + (x_2 - x_1) \vee 0 \| \leqq \| x_1 + x_2 \| = \| x \| \leqq 1$ and $f(x') - g(x') = -f(x_1) + f(x_2) + g(x_1) - g(x_2) > f(x_1) + f(x_2) + g(x_1) + g(x_2) - 2\epsilon = f(x) + g(x) - 2\epsilon > \| f + g \| - 3\epsilon$. Hence we have $\| f - g \| > \| f + g \| - 3\epsilon$, and since $\epsilon > 0$ is arbitrary, we must have $\| f - g \| \geqq \| f + g \|$.

In order to prove the inverse inequality, let $\epsilon > 0$ be an arbitrary positive number. Then there exists an $x \epsilon E$ such that $\| x \| \leqq 1$ and $f(x) - g(x) > \| f - g \| - \epsilon$. Clearly $x' = x \vee 0 + (-x) \vee 0$ satisfies $\| x' \| = \| x \| \leqq 1$ (by (1.15)) and $f(x') + g(x') = f(x \vee 0) + f((-x) \vee 0) + g(x \vee 0) + g((-x) \vee 0) \geqq f(x \vee 0) - f((-x) \vee 0) - g(x \vee 0) + g((-x) \vee 0) = f(x) - g(x) > \| f - g \| - \epsilon$. Hence we have $\| f + g \| > \| f - g \| - \epsilon$, and since $\epsilon > 0$ is arbitrary, we must have $\| f + g \| \geqq \| f - g \|$.

THEOREM 15. *The conjugate space $(AM)^*$ of an abstract (M)-space (AM) is an abstract (L)-space, and conversely the conjugate space $(AL)^*$ of an abstract (L)-space (AL) is an abstract (M)-space.*

PROOF. By Lemma 16.1, we need only prove the following two facts:

$$(16.2) \quad f, g \,\epsilon\, (AM)^*, \quad f \geqq 0, \quad g \geqq 0 \quad \textit{imply} \quad \| f + g \| = \| f \| + \| g \|,$$

$$(16.3) \quad f, g \,\epsilon\, (AL)^*, \quad f \geqq 0, \quad g \geqq 0 \quad \textit{imply} \quad \| f \vee g \| = \max (\| f \|, \| g \|).$$

To prove (16.2), let $f, g \,\epsilon\, (AM)^*, f \geqq 0, g \geqq 0$, and let $\epsilon > 0$ be an arbitrary positive number. Then there exist $x, y \,\epsilon\, (AM)$ such that $\| x \| \leqq 1, \| y \| \leqq 1$, $f(x) > \| f \| - \epsilon$ and $g(y) > \| g \| - \epsilon$. We may assume $x \geqq 0, y \geqq 0$, since otherwise we can replace x and y by $x \vee 0$ and $y \vee 0$. Then $z = x \vee y$ satisfies $\| z \| = \max(\| x \|, \| y \|) \leqq 1$ and $f(z) + g(z) \geqq f(x) + g(y) > \| f \| + \| g \| - 2\epsilon$. Hence $\| f + g \| > \| f \| + \| g \| - 2\epsilon$, and since $\epsilon > 0$ is arbitrary, we have $\| f + g \| \geqq \| f \| + \| g \|$. The inverse inequality $\| f + g \| \leqq \| f \| + \| g \|$ is clear.

Next, in order to prove (16.3), let us first prove that there exists a unit element f_0 (in the sense of section 3) in $(AL)^*$. Indeed, if we put

$$(16.4) \qquad f_0(x) = \| x \vee 0 \| - \| (-x) \vee 0 \|$$

for any $x \,\epsilon\, (AL)$, then we have

$$(16.5) \qquad | f_0(x) | \leqq \| x \|,$$

$$(16.6) \qquad f_0(x + y) = f_0(x) + f_0(y),$$

$$(16.7) \qquad f_0(\lambda x) = \lambda f_0(x),$$

$$(16.8) \qquad f \,\epsilon\, (AL)^*, \qquad \| f \| \leqq 1 \quad \textit{imply} \quad f \leqq f_0 .$$

[14] $k = f \wedge g$ is defined, for $x \geqq 0$, by $k(x) = \text{g.l.b.} \{f(x_1) + g(x_2)\}$, where g.l.b. means the greatest lower bound for all decompositions of x: $x = x_1 + x_2$, $x_1 \geqq 0$, $x_2 \geqq 0$.

(16.5) and (16.7) are clear; (16.6) follows from the relation

$$\|(x+y) \vee 0\| + \|(-x) \vee 0\| + \|(-y) \vee 0\|$$
$$= \|(-x-y) \vee 0\| + \|x \vee 0\| + \|y \vee 0\|,$$

which is a direct consequence of (16.1) and the equality

$$(x+y) \vee 0 + (-x) \vee 0 + (-y) \vee 0 = (-x-y) \vee 0 + x \vee 0 + y \vee 0.$$

(16.8) is also clear, since $\|f\| \leqq 1$, $x \geqq 0$ imply $f(x) \leqq \|x\| = \|x \vee 0\| - \|(-x) \vee 0\| = f_0(x)$.

Thus we have proved that $f_0(x)$ is a unit element of $(AL)^*$ in the sense of section 3. Hence, by Remark 1 of section 3, $(AL)^*$ is an abstract (M)-space. The proof of Theorem 15 is completed.

EXAMPLE 1. If $(AM) = C(\Omega)$, where Ω is a compact Hausdorff space, then $(AM)^* = \mathfrak{M}(\Omega)$ (see Theorem 10). $\mathfrak{M}(\Omega)$ is clearly an abstract (L)-space. For further examples, see Remark after Theorem 10.

EXAMPLE 2. If $(AL) = L(\Omega; m)$, where Ω is a compact Hausdorff space with a completely additive measure $m(E)$ such that $m(\Omega) = 1$ (see Theorem 14), then $(AL)^* = M(\Omega; m)$ (for the definition of $M(\Omega; m)$ see section 2). The same is also true if $(AL) = L(\Omega; m) = \sum_{\alpha \in \mathfrak{M}} L(\Omega_\alpha; m)$, where $\Omega = \sum_{\alpha \in \mathfrak{M}} \Omega_\alpha$ and each Ω_α is compact.

REMARK 1. Theorem 15 does not mean that (AM) and (AL) are reflexive. In fact, these are reflexive only when they are finite dimensional.

REMARK 2. For any abstract (M)-space (AM), its second conjugate space $(AM)^{**}$ is also an abstract (M)-space (by using Theorem 15 twice) which contains (AM) as a closed linear subspace.[15] The same argument is true also for

[15] We should be careful about the following situation: (AM) is embedded isometrically into $(AM)^{**}$ and this embedding is partial order preserving, i.e. if $x, y \in (AM)$ correspond to $X, Y \in (AM)^{**}$, then $x \geqq y$ implies $X \geqq Y$. But it is not obvious that in this case $x \vee y$ corresponds to $X \vee Y$. In order to show this, let us put $z = x \vee y$, $Z = X \vee Y$, and let Z' be the element of $(AM)^{**}$ to which z corresponds. It is then clear that we have $Z' \geqq Z$ (this is a direct consequence of $Z' \geqq X$ and $Z' \geqq Y$). In order to prove the inverse inequality, let us remember that, for $f \geqq 0$, $Z(f)$ is defined by $Z(f) = \text{l.u.b.}\ \{X(f_1) + Y(f_2)\} = \text{l.u.b.}\ \{f_1(x) + f_2(y)\}$, where l.u.b. means the least upper bound for all decompositions of f: $f = f_1 + f_2$, $f_1 \geqq 0$, $f_2 \geqq 0$. On the other hand, if we represent (AM) concretely as a subspace of the space $C(\Omega)$ of all real valued continuous functions defined on a compact Hausdorff space Ω (Theorem 1), then $z(t) = \max\ (x(t), y(t))$ for all $t \in \Omega$, and every $f \in (AM)^*$ with $f \geqq 0$ may be considered as a regular measure $\mu(E)$ defined for all Borel sets E of Ω (Theorem 9). Let $O \subseteq \Omega$ be the set of all t at which $x(t) > y(t)$. Then O is an open set which is at the same time an F_σ. Let us now define two measures $\mu_1(E)$ and $\mu_2(E)$ by $\mu_1(E) = \mu(EO)$ and $\mu_2(E) = \mu(E - EO)$. Then $\mu_1(E)$ and $\mu_2(E)$ are both regular measures defined on Borel sets of Ω, and if we denote by f_1 and f_2 the corresponding elements of $(AM)^*$, then we have $f = f_1 + f_2$, $f_1 \geqq 0$, $f_2 \geqq 0$ and it is not difficult to see that $f(z) = f_1(x) + f_2(y)$. Hence $Z(f) \geqq f(z) = Z'(f)$ for every $f \geqq 0$, i.e. $Z \geqq Z'$. Thus we proved the inverse inequality, and hence $Z = Z'$.

(AL) and $(AL)^{**}$.[16] It is, however, to be noted that $(AM)^{**}$ is always an abstract (M)-space with a unit element, while this is not the case for $(AL)^{**}$ even if (AL) has an F-unit.

Hence, every abstract (M)-space can be embedded in an abstract (M)-space with a unit element. Consequently, in discussing the realization of abstract (M)-spaces, we have only to treat the case when there exists a unit element. This fact explains the relation between Theorem 1 and Theorem 2.

REMARK 3. Let (AL) be an abstract (L)-space. Then $(AL)^*$ is an abstract (M)-space with a unit element. Hence, by Theorem 2, there exists a compact Hausdorff space Ω such that $(AL)^*$ is isometric and lattice isomorphic to $C(\Omega)$. Hence every element $x \in (AL)$ can be considered as a bounded linear functional $x(f)$ defined on $C(\Omega)$. Consequently, by Theorem 10, $x(f)$ can be represented as an integral:

$$x(f) = \int_\Omega f(t)\mu(dt) \tag{16.9}$$

where $\mu(E)$ is a completely additive regular real-valued set function defined for all Borel sets E of Ω. From this we have

THEOREM 16. *For any abstract (L)-space (AL), there exists a compact Hausdorff space Ω such that (AL) is isometric and lattice isomorphic to a closed linear subspace of the space $\mathfrak{M}(\Omega)$ of all completely additive regular real-valued set-functions $\mu(E)$ defined for all Borel sets E of Ω (where norm and partial ordering are defined as in Theorem* 10).

In the same manner we can prove (see Theorem 14, Remark 2 after Theorem 14, and Example 2 after Theorem 15) the following

THEOREM 17. *For any abstract (M)-space (AM), there exists a Hausdorff space $\Omega = \sum_{\alpha\in\mathfrak{M}} \Omega_\alpha$, where each Ω_α is compact, and a completely additive measure defined for all Borel sets of each Ω_α such that (AM) is isometric and lattice isomorphic to a closed linear subspace of the space $M(\Omega; m) = \sum_{\alpha\in\mathfrak{M}} M(\Omega_\alpha; \mathfrak{M})$ of all bounded measurable functions $x(t)$ defined on Ω, where we call $x(t)$ measurable if and only if $x(t)$ is measurable on each Ω_α. (For the definition of $M(\Omega; m)$ see section* 2.)

Theorems 16 and 17 are weaker than Theorems 14 and 1 (or 2) respectively.

INSTITUTE FOR ADVANCED STUDY

BIBLIOGRAPHY

(1) G. Birkhoff, *Dependent probabilities and the space* (L), Proc. Nat. Acad. Sci., U.S.A., **24** (1938), 154–159.

[16] It is also not difficult to see that the embedding of (AL) into $(AL)^{**}$ is not only isometric and partial order preserving but also lattice isomorphic. We have only to consider the concrete representation of (AL) and the general form of positive linear functionals defined on this concrete (AL) (see Example 2 above). The arguments of the same kind as in footnote 15 will show that, if $x, y \in (AL)$ correspond to $X, Y \in (AL)^{**}$, then $z = x \vee y$ corresponds to $Z = X \vee Y$. Recently it was proved by Dr. Alaoglu that when we embed an arbitrary Banach lattice E into its second conjugate E^{**} (in the ordinary way) this embedding is always a lattice isomorphism.

(2) G. Birkhoff, *Lattice theory*, Amer. Math. Soc. Colloquium Publications, 1940.
(3) H. F. Bohnenblust, Characterization of L_p-spaces, Duke Math. Jour., **6** (1940), 627–640.
(4) E. Čech, *On bicompact spaces*, Annals of Math., **38** (1937), 823–844.
(5) H. Freudenthal, *Teilweise geordnete Moduln*, Proc. Acad. Amsterdam, **39** (1936), 641–651.
(6) I. Gelfand, *On normed rings*, C. R. URSS, **23** (1939), 430–432.
(7) I. Gelfand and A. Kolmogoroff, *On rings of continuous functions of topological spaces*, C. R. URSS, **22** (1939), 11–15.
(8) S. Kakutani, *Two fixed point theorems concerning bicompact convex sets*, Proc. Imp. Acad. Japan, **14** (1938), 242–245.
(9) S. Kakutani, *Mean ergodic theorem in abstract (L)-spaces*, Proc. Imp. Acad. Japan, **15** (1939), 121–123.
(10) S. Kakutani, *Weak topology, bicompact sets and the principle of duality*, Proc. Imp. Acad. Japan, **16** (1940), 63–67.
(11) S. Kakutani, *Concrete representation of abstract (L)-spaces and the mean ergodic theorem*, Annals of Math., **42** (1941), 523–537.
(12) L. Kantorovitch, *Lineare halbgeordnete Räume*, Recueil Math., **44** (1937), 121–165.
(13) J. von Neumann, *On an algebraic generalization of quantum mechanical formalism*, Part 1, Recueil Math., **43** (1936), 415–484.
(14) J. von Neumann, *Invariant measure*, Princeton lectures, 1940–41.
(15) S. Saks, *Integration in abstract metric spaces*, Duke Math. Jour., **4** (1938), 408–411.
(16) G. Šilov, *Ideals and subrings of the ring of continuous functions*, C. R. URSS, **22** (1939), 7–10.
(17) M. H. Stone, *Applications of the theory of Boolean rings to general topology*, Trans. Amer. Math. Soc., **41** (1937), 376–381.
(18) M. H. Stone, *A general theory of spectra*, Proc. Nat. Acad. Sci., U. S. A., **26** (1940), 280–283.
(19) K. Yosida, *On the theory of spectra*, Proc. Imp. Acad. Japan, **16** (1940) 578–583.

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

CONCRETE REPRESENTATION OF (M)-SPACES

BY H. F. BOHNENBLUST AND S. KAKUTANI

(Received January 20, 1941)

1. Introduction

In the preceding paper[1] one of us proves that any Banach lattice which satisfies the condition

$$(1.1) \qquad 0 < x, \qquad 0 < y \text{ implies } \| x \vee y \| = \text{Max } (\| x \|, \| y \|)$$

can be represented as a sublattice of the lattice of all real valued continuous functions over a compact[2] Hausdorff space Ω. The purpose of this note is to obtain the same result from the weaker condition

$$(1.2) \qquad x \wedge y = 0 \text{ implies } \| x \vee y \| = \text{Max } (\| x \|, \| y \|).$$

Our proof yields, in particular, the result that in a Banach lattice (1, 2) implies (1, 1). It does not seem possible, however, to establish this implication directly. Our method consists in obtaining the representation of the lattice independently of the previous result and then (1.1) is trivially true for continuous functions.

If the additional assumption is made that the Banach lattice contains a unit element, it is shown in the preceding paper that the representation uses *all* continuous functions. In this form the result was obtained independently by Mark and Selim Krein.[3] The method we shall use in the present paper is an adaptation of that followed in Krein's paper.

2

Let E be a Banach lattice, i.e.

$$(2.1) \qquad E \text{ is a Banach space,}$$

$$(2.2) \qquad E \text{ is a vector lattice,}$$

$$(2.3) \qquad 0 < x < y \text{ implies } \| x \| \leqq \| y \|,^{4}$$

$$(2.4) \qquad x \wedge y = 0 \text{ implies } \| x - y \| = \| x + y \|.$$

---

[1] S. Kakutani, *Concrete representation of abstract* (M)*-spaces*. Annals of Math., 42 (1941), pp. 994–1024.

[2] We use the term "compact" as "open coverings can be reduced to finite coverings."

[3] M. and S. Krein, *On an inner characteristic of the set of all continuous functions on a bicompact Hausdorff space*, Comptes Rendus U.R.S.S., 27 (1940), 427–430.

[4] It is interesting to notice that condition (2, 3) cannot be deduced from (2.4) and (2.5) combined.

Assume furthermore that the condition

(2.5) $$x \wedge y = 0 \text{ implies } \|x + y\| = \text{Max}(\|x\|, \|y\|)$$

holds in E.

The conjugate space E^* is also a Banach lattice. Denote by S^* its unit sphere ($\|f\| \leqq 1$) and by S^*_+ the positive part ($f \geqq 0$) of S^*. The set S^*_+ is *bounded, convex* and *w^*-closed.*[5] The extreme points of S^*_+ form a set which will be denoted by T. By a theorem of Krein and Milman[6] we have

THEOREM 2.1. *The set S^*_+ is the smallest convex, w^*-closed set which contains T.*

THEOREM 2.2. *With the exception of* 0, *every element of T has norm* 1.

PROOF: If $f \neq 0$ belongs to T, then $f/\|f\|$ belongs to S^*_+ and the element f is an *interior* point of the segment joining 0 and $f/\|f\|$ unless $\|f\| = 1$.

THEOREM 2.3. *If $f \in T$ and if $u, v \in E$, $u \wedge v = 0$ then $f(u)\cdot f(v) = 0$.*

The proof is by contradiction. Assume $f \in T, f(u) > 0, f(v) > 0$ and $u \wedge v = 0$. Two elements of $g, h \in E^*$ will be constructed such that

(2.6)
$$\begin{array}{ll} \text{i)}\ \ 0 \leqq g \leqq f & \text{ii)}\ \ g(v) = 0 \\ \text{iii)}\ \ g(u) = f(u) > 0 & \text{iv)}\ \ f = g + h \\ \multicolumn{2}{c}{\text{v)}\ \ 1 = \|f\| = \|g\| + \|h\|.} \end{array}$$

Conditions i) and iii) show that $0 < h$. Thus $g/\|g\|$ and $h/\|h\|$ belong to S^*_+ and

$$f = \|g\| \cdot \frac{g}{\|g\|} + h \cdot \frac{h}{\|h\|}$$

proves that f is an interior point of the segment joining $g/\|g\|$ and $h/\|h\|$. This is a contradiction to the assumption $f \in T$ since $g/\|g\| \neq h/\|h\|$ by ii) and iii).

Definition of g. For $x \geqq 0$ put

$$g(x) = \sup_n f(nu \wedge x)$$

and for $x = x^+ - x^-$, $g(x) = g(x^+) - g(x^-)$. We omit the details showing that $g \in E^*$ and that i), ii), iii) are satisfied.

Definition of h: $h = f - g$. Condition iv) is true and to prove the last one, v), we need only show that

$$\|g\| + \|h\| \leqq 1 + 4\epsilon$$

for any $\epsilon > 0$. Choose $\epsilon > 0$ and determine x_1 such that $\|x_1\| = 1$ and $g(x_1) > \|g\| - \epsilon$. We may assume $x_1 > 0$ since otherwise it could be replaced by $|x_1|$.

[5] w^*-topology in E^* is the weak topology with reference to the elements of E.

[6] M. Krein and D. Milman, *On extreme points of regular convex sets*, Studia Math., IX (1940), 133–138.

By the definition of g there will exist a positive integer n_1 for which

$$f(n_1 u \wedge x_1) > g(x_1) - \epsilon > \| g \| - 2\epsilon.$$

Put $x_2 = x_1 \wedge n_1 \cdot u$; then

$$\| x_2 \| \leqq 1, \qquad g(x_2) = f(x_2) > \| g \| - 2\epsilon, \qquad h(x_2) = 0. \tag{2.7}$$

Now we turn our attention to h. Determine an element y_1 such that

$$0 \leqq y_1, \qquad \| y_1 \| = 1, \qquad h(y_1) > \| h \| - \epsilon,$$

and choose n_2 to guarantee $1 < \epsilon \cdot (n_2 + 1)$. Put

$$y_2 = (y_1 - n_2 \cdot x_2) \vee 0.$$

Obviously $0 \leqq y_2 \leqq y_1$ and hence $\| y_2 \| \leqq 1$. Also

$$h(y_2) = h(y_1 \vee n_2 x_2) - n_2 h(x_2) = h(y_1 \vee n_2 x_2) \geqq h(y_1) > \| h \| - \epsilon. \tag{2.8}$$

We evaluate the norm of $x_2 \wedge y_2$. Notice that

$$x_2 \wedge y_2 = 0 \vee [x_2 \wedge (y_1 - n_2 \cdot x_2)].$$

But

$$n_2[x_2 \wedge (y_1 - n_2 x_2)] \leqq n_2 x_2$$
$$[x_2 \wedge (y_1 - n_2 x_2)] \leqq y_1 - n_2 x_2$$

and thus by addition

$$(n_2 + 1)[x_2 \wedge (y_1 - n_2 x_2)] \leqq y_1 .$$

This implies $0 \leqq x_2 \wedge y_2 \leqq y_1/(n_2 + 1) \leqq \epsilon \cdot y_1$ and therefore

$$\| x_2 \wedge y_2 \| \leqq \epsilon. \tag{2.9}$$

Finally let us put

$$x_3 = x_2 - (x_2 \wedge y_2) \qquad \text{and} \qquad y_3 = y_2 - (x_2 \wedge y_2).$$

Evidently $x_3 \wedge y_3 = 0$, $\| x_3 \| \leqq 1$, $\| y_3 \| \leqq 1$, $\| x_3 + y_3 \| = \text{Max}\,(\| x_3 \|, \| y_3 \|) \leqq 1$. Since $0 \leqq h(x_2 \wedge y_2) \leqq h(x_2) = 0$ it follows from (2.8) that

$$h(y_3) = h(y_2) > \| h \| - \epsilon.$$

On the other hand, by (2.7) and (2.9)

$$g(x_3) = g(x_2) - g(x_2 \wedge y_2) > \| g \| - 2\epsilon - \| g \| \cdot \epsilon \geqq \| g \| - 3\epsilon.$$

Combining these results we obtain

$$\begin{aligned} 1 \geqq f(x_3 + y_3) &= g(x_3) + g(y_3) + h(x_3) + h(y_3) \\ &> \| g \| - 3 \cdot \epsilon + 0 + 0 + \| h \| - \epsilon = \| g \| + \| h \| - 4\epsilon. \end{aligned}$$

The proof of v) is complete and theorem 2.3 is established.

THEOREM 2.4. *Let Ω be the w^*-closure of T. If $f \in \Omega$ then $f \geqq 0$ and if u, $v \in E$, $u \wedge v = 0$, then $f(u) \cdot f(v) = 0$.*

PROOF: For any $\epsilon > 0$ there exists $g \in T$ with

$$| g(u) - f(u) | < \epsilon,$$

$$| g(v) - f(v) | < \epsilon.$$

By theorem 2.3 either $g(u)$ or $g(v) = 0$. Hence either $|f(u)|$ or $|f(v)| < \epsilon$, which implies of course $f(u) \cdot f(v) = 0$. The proof of $f \geqq 0$ is even simpler.

3. The Representation

Each element x of E determines a function φ_x defined over Ω by $\varphi_x(f) = f(x)$. The set Ω will be considered as a topological space by taking the w^*-topology of E^*. It is a Hausdorff space and since S^* is w^*-compact, Ω will also be w^*-compact. In the space Ω the function φ_x is continuous. Furthermore we have evidently

(3.1) $$\varphi_{\lambda x + \mu y} = \lambda \varphi_x + \mu \varphi_y$$

(3.2) $$x \geqq 0 \text{ implies } \varphi_x \geqq 0.$$

[The last inequality means: for any $f \in \Omega$, $\varphi_x(f) \geqq 0$].

If $x_1 \wedge x_2 = 0$ then $f(x_1) \cdot f(x_2) = 0$ for any $f \in \Omega$ by theorem 2.4. Thus $\varphi_{x_1}(f) \cdot \varphi_{x_2}(f) = 0$ and therefore $\varphi_{x_1} \wedge \varphi_{x_2} = 0$. For any x_1 and x_2, the elements $x_1 - (x_1 \wedge x_2)$ and $x_2 - (x_1 \wedge x_2)$ are disjoint. We have then

$$(\varphi_{x_1} - \varphi_{x_1 \wedge x_2}) \wedge (\varphi_{x_2} - \varphi_{x_1 \wedge x_2}) = 0$$

or

(3.3) $$\varphi_{x_1} \wedge \varphi_{x_2} = \varphi_{x_1 \wedge x_2} \quad \text{and also} \quad \varphi_{x_1} \vee \varphi_{x_2} = \varphi_{x_1 \vee x_2}$$

Finally, if we define $\| \varphi_x \| = \max | \varphi_x(f) |$ $(f \in \Omega)$, we prove

(3.4) $$\| x \| = \| \varphi_x \|.$$

It is sufficient to prove (3.5) for positive x, since $\| \, | x | \, \| = \| x \|$ and $\| \varphi_{|x|} \| = \| \varphi_x \|$. Evidently $\| \varphi_x \| \leqq \| x \|$. Let $f_0 \in E^*$, $\| f_0 \| = 1$, $f_0(x) = 1$. We may assume $f_0 \geqq 0$ since otherwise $| f_0 |$ could be taken. If now $\| \varphi_x \| = \| x \| - \epsilon$, $\epsilon > 0$ then

$$f \in \Omega \rightarrow f(x) \leqq \| x \| - \epsilon,$$

thus

$$f \in S^*_+ \quad f(x) \leqq \| x \| - \epsilon$$

and we obtain a contradiction for f_0.

The equation (3.4) shows in particular that $\varphi_x = 0$ implies $x = 0$, in other words: to different x's correspond different φ_x's.

The relations (3.1), (3.2), (3.3) and (3.4) give us the desired representation.

PRINCETON, N. J.

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

SELECTIVE EQUATIONS

BY E. T. BELL

(Received October 26, 1940)

A detail in the systematic solution of certain compound multiplicative diophantine systems[1] leads to a new type of completely solvable diophantine problem which may be of independent interest, as but few such problems are known.

1. Selective operators and systems

If $n_1, \cdots, n_s$ are integers $\geqq 0$, and if $s > 1$, $(n_1, \cdots, n_s)'$ denotes the least, and $[n_1, \cdots, n_s]'$ the greatest, of $n_1, \cdots, n_s$ if at least two of these integers are unequal; if $n_1 = \cdots n_s = n$, the value of each of these symbols is n. By definition, if m is an integer $\geqq 0$, $(m)' = [m]' = m$. Hence, for $s \geqq 1$,

$$[(n_1, \cdots, n_s)']' = (n_1, \cdots, n_s)' = ((n_1, \cdots, n_s)')',$$

$$([n_1, \cdots, n_s]')' = [n_1, \cdots, n_s]' = [[n_1, \cdots, n_s]']'.$$

The symbols $()'$, $[]'$ will be called simple selective operators.

Let each symbol $\{\}'$ denote a definite one of $()'$, $[]'$. In $\{n_1, \cdots, n_s\}'$ replace n_j by $\{n_{j_1}, \cdots, n_{j_{s_j}}\}'$, where the n_{j_i} are integers $\geqq 0$. In the result repeat the process, and so on, a finite number of times. The symbol $\{\cdots\{\cdots\{\cdots\}'\cdots\}'\cdots\}'$ so obtained will be called a compound selective operator. It will be seen that compound operators may be replaced by simple operators in the type of equations to be defined.

If the results of operating on the set $n_1, \cdots, n_s$ with two selective operators are equal for all sets $n_1, \cdots, n_s$, we say that the operators are identical; otherwise, the operators are distinct. The result of operating on the ordered set $n_1, \cdots, n_s$ with the selective operator σ' will be written $\sigma'(n_1, \cdots, n_s)$.

A selective system is defined as follows. Let $\sigma_i', \cdots, \sigma_j', \cdots, \sigma_k', \cdots, \sigma_l'$ be any selective operators (simple or compound, and unrestricted with respect to distinctness). The $a, \alpha, a_i, \cdots, b, \beta, b_i, \cdots, c, \gamma, c_k, \cdots, d, \delta, d_l$ are constant integers > 0, and all the x's, $\cdots$, y's, $\cdots$, z's, $\cdots$, w's are variable integers $\geqq 0$. The $A, \cdots, B, \cdots, C, \cdots, D$ are constant integers $\geqq 0$. If any of the A's are zero, the terms involving them are suppressed, and likewise for the B's, $\cdots$, C's, $\cdots$, D's. We write

$$\sigma_i'(x_{i,1}, \cdots, x_{i,m_i}) \equiv X_i', \cdots, \sigma_j'(y_{j,1}, \cdots, y_{j,n_j}) \equiv Y_j',$$

$$\cdots \qquad \cdots \qquad \cdots$$

$$\sigma_k'(z_{k,1}, \cdots, z_{k,r_k}) \equiv Z_k', \cdots, \sigma_l'(w_{l,1}, \cdots, w_{l,s_l}) \equiv W_l'.$$

[1] E. T. Bell, Proc. Nat. Acad. Sci., 26 (1940), 462–466.

A selective system is then a system of equations of the type

$$\sum_{i=1}^{a} A_i x_i + \sum_{i=1}^{\alpha} a_i X_i' = \cdots = \sum_{j=1}^{b} B_j y_j + \sum_{j=1}^{\beta} b_j Y_j',$$

$$\cdots \qquad \cdots \qquad \cdots$$

$$\sum_{k=1}^{c} C_k z_k + \sum_{k=1}^{\gamma} c_k Z_k' = \cdots = \sum_{l=1}^{d} D_l w_l + \sum_{l=1}^{\delta} d_l W_l',$$

in which the total number of equality signs is finite, and in which at least two sets of variables in different rows have at least one variable in common, while in any one row no two sets of variables are identical, but any two sets in that row may have variables in common.

It is required to find all integer values $\geqq 0$ of the variables satisfying the system. We shall say that this problem is solved when a method is prescribed for exhibiting the required integers in any given selective system in a finite number of non-tentative steps. The solution is in terms of a certain minimum number N of parameters ranging independently over all integers $\geqq 0$. The above definition of 'solution' is inserted because for even apparently trivial systems the N necessary and sufficient for a complete solution may be so large that it is impossible to write out the solution. In fact, if N' is any given positive integer, so elementary a selective system as

$$a(x_1, x_2)' + a(y_1, y_2)' = b(x_1, z_2)' + b(w_1, w_2)'$$

can be constructed for which $N > N'$. If in this example $N' = 1{,}000{,}000$, it suffices to take $a = 1000$, $b = 1001$.

A selective system is thus one of quadruple composition, the four operations involved being addition and multiplication of non-negative integers, and $()'$, $[]'$.

2. Dual of a selective system

As might be anticipated from the definitions in §1 and a usual method for finding the G. C. D. and L. C. M. of a set of integers from the canonical decompositions (into products of powers of distinct primes) of the integers concerned, a selective system may be restated in terms of G. C. D.'s, L. C. M.'s, and products instead of sums. It will be seen that the solution of either system implies that of the other. The connection between the two types is the solution of G. C. D.-L. C. M. equations discussed in §§3, 4.

With $n_1, \cdots, n_s$ as in §1, $(n_1, \cdots, n_s)$ denotes the G. C. D., and $[n_1, \cdots, n_s]$ the L. C. M., of $n_1, \cdots, n_s$; $\{\}$ denotes a definite one of $()$, $[]$. The rest of §1, down to the selective system exhibited, may be rewritten with $\{\}$, σ in place

of $\{\}'$, σ', and X_i, Y_j, Z_k, W_l in place of X'_i, Y'_j, Z'_k, W'_l. The system is replaced by

$$\prod_{i=1}^{a} x_i^{A_i}\cdot\prod_{i=1}^{\alpha} X_i^{a_i} = \cdots = \prod_{j=1}^{b} y_j^{B_j}\cdot\prod_{j=1}^{\beta} Y_j^{b_j},$$

$$\cdots \qquad \cdots \qquad \cdots$$

$$\prod_{k=1}^{c} z_k^{C_k}\cdot\prod_{k=1}^{\gamma} Z_k^{c_k} = \cdots = \prod_{l=1}^{d} w_l^{D_l}\cdot\prod_{l=1}^{\delta} W_l^{d_l};$$

and the rest of §1 follows with a few obvious verbal changes. This system and that in §1 will be called duals of each other, for the following reasons.

It will be shown that the complete solution of the above system is of the form

$$\begin{aligned}
x_i &= \prod\theta_s^{x_{i,s}}(i = 1, \cdots, a),\\
x_{i,t} &= \prod\theta_s^{x_{i,t,s}}(i = 1, \cdots, \alpha; t = 1, \cdots, m_i);\\
&\cdots \qquad \cdots \qquad \cdots\\
y_j &= \prod\theta_s^{y_{j,s}}(j = 1, \cdots, b),\\
y_{j,t} &= \prod\theta_s^{y_{j,t,s}}(j = 1, \cdots, \beta; t = 1, \cdots, n_j);\\
&\cdots \qquad \cdots \qquad \cdots\\
z_k &= \prod\theta_s^{z_{k,s}}(k = 1, \cdots, c),\\
z_{k,t} &= \prod\theta_s^{z_{k,t,s}}(k = 1, \cdots, \gamma; t = 1, \cdots, r_k);\\
&\cdots \qquad \cdots \qquad \cdots\\
w_l &= \prod\theta_s^{w_{l,s}}(l = 1, \cdots, d),\\
w_{l,t} &= \prod\theta_s^{w_{l,t,s}}(l = 1, \cdots, \delta; t = 1, \cdots, s_l);
\end{aligned}$$

in which the product refers to $s = 1, \cdots, N$, the θ's are integer parameters, all exponents are constant integers $\geqq 0$ and the parameters are subject to a certain finite set of conditions, called the G. C. D. conditions,[2] of the form

$$1 = \left(\prod_{i=1}^{m}\theta_{a_i}, \cdots, \prod_{i=1}^{n}\theta_{b_i}\right) = \cdots = \left(\prod_{i=1}^{u}\theta_{c_i}, \cdots, \prod_{i=1}^{v}\theta_{d_i}\right).$$

From this, the solution of the selective system in §1 is written down by replacing all multiplications by additions, and restricting the parameters $\theta_1, \cdots, \theta_N$ to range over only integers $\geqq 0$:

$$x_i = \sum_{s=1}^{N} x_{i,s}\theta_s \qquad (i = 1, \cdots, a),$$

$$x_{i,t} = \sum_{s=1}^{N} x_{i,t,s}\theta_s \quad (i = 1, \cdots, \alpha; t = 1, \cdots, m_i);$$

---

[2] E. T. Bell, Amer. Journ. Math., 55 (1933), 50–66.

and similarly for the rest. The G. C. D. conditions are correspondingly replaced by

$$0 = \left(\sum_{i=1}^{m} \theta_{a_i}, \cdots, \sum_{i=1}^{n} \theta_{b_i}\right)' = \cdots = \left(\sum_{i=1}^{u} \theta_{c_i}, \cdots, \sum_{i=1}^{v} \theta_{d_i}\right)'.$$

Each of these conditions is of the form

$$0 = (A_1' + \cdots + A_f', \cdots, B_1' + \cdots + B_g')',$$

in which the A', B' are integers $\geqq 0$. From the definition of $()'$, it follows that this condition is satisfied if and only if at least one of the following sets of conditions is satisfied,

$$\begin{gathered} A_1' = \cdots = A_f' = 0, \\ \cdots \qquad \cdots \\ B_1' = \cdots = B_g' = 0; \end{gathered}$$

and similarly for all in the original set. The complete solution of the selective system may therefore be separated into sets of solutions, all the solutions in a particular set being obtained from the general solution by suppressing those parameters which occur in a selected subset of conditions sufficient to satisfy the complete set of conditions; the remaining parameters in the solution are then integers $\geqq 0$ subject to no conditions.

A detailed proof of the duality described is quite simple, but leads to somewhat complicated formulas. It will be sufficient to indicate how the proof may be written out in detail.

Let $\qquad p_1 = 2, \qquad p_2 = 3, \qquad p_3 = 5, \qquad p_4 = 7, \cdots$

be the natural primes in ascending order. The canonical decompositions of the positive integers n, n_i, $n_{i,j}$, $\cdots$ are written

$$n = \prod p_\xi^{n_\xi}, \qquad n_i = \prod p_\xi^{n_{i,\xi}}, \qquad n_{i,j} = \prod p_\xi^{n_{i,j,\xi}}, \cdots,$$

where the product refers to $\xi = 1, 2, 3, \cdots$. (In other connections, this device is useful for absolute constants; for example, $10 = p_1^{10_1} p_2^{10_2} p_3^{10_3} \cdots$, so that $10_1 = 1$, $10_2 = 0$, $10_3 = 1$, $10_s = 0$, $s > 3$.) In each decomposition there is a finite index ξ' such that $\xi = 0$ if $\xi \geqq \xi'$, $\xi > 0$ if $\xi < \xi'$. This notational device is applied to the variables of the system in this section, when it follows that the system is equivalent to the selective system in §1. For if $\prod p_\xi^{f_\xi} = \prod p_\xi^{g_\xi}$, then $f_\xi = g_\xi$ for all integers $\xi > 0$; also,

$$\begin{aligned} \{u_1, \cdots, u_n\}^m &= \{\prod p_\xi^{u_{1,\xi}}, \cdots, \prod p_\xi^{u_{n,\xi}}\}^m, \\ &= \prod p_\xi^{m\{u_{1,\xi}, \cdots, u_{n,\xi}\}'}, \end{aligned}$$

by the definitions of $\{\}$, $\{\}'$. The parameters θ in the solution are replaced by their canonical decompositions, $\theta_j = \prod p_\xi^{\theta_{j,\xi}}$; whence the rule for obtaining the solution of the dual selective system is immediate. It is not necessary to include

the signs of the numbers generated as the θ's in the solution of the $\{\}$-equation range over all integers; but if desired this may be done by noting that for any integer c, $c = \operatorname{sgn} c \cdot \prod p_\xi^{c_\xi}$, where sgn has its usual meaning.

3. G. C. D.-L. C. M. equations

The solution of a selective system is obtained as indicated in §2 from that of the dual system. Obviously, the solution of the dual system may be found by repeated applications of the solution of

$$u_1^{a_1} \cdots u_r^{a_r} \{x_1, \cdots, x_n\}_1^c = v_1^{b_1} \cdots v_s^{b_s} \{y_1, \cdots, y_m\}_2^d,$$

in which $\{\}_1$, $\{\}_2$ are unrestricted (identical or distinct, simple or compound) symbols constructed from $()$, $[]$ in the same way as $\{\}'$ from $()'$, $[]'$ in §1; the a, b are constant integers $\geqq 0$, and c, d integers > 0; and the u, x, v, y are variable integers. Introducing new variables x_{n+1}, y_{m+1}, we first solve

$$\{x_1, \cdots, x_n\}_1 = x_{n+1}, \qquad \{y_1, \cdots, y_m\}_2 = y_{m+1},$$

as described presently. Each solution is by power products in a certain minimum number of integer parameters, subject to a set of G. C. D. conditions; say this gives

$$x_{n+1} = \phi_1^{\alpha_1} \cdots \phi_f^{\alpha_f}, \qquad y_{n+1} = \psi_1^{\beta_1} \cdots \psi_g^{\beta_g}.$$

The equation then becomes

$$u_1^{a_1} \cdots u_r^{a_r} \phi_1^{c\alpha_1} \cdots \phi_f^{c\alpha_f} = v_1^{b_1} \cdots v_s^{b_s} \psi_1^{d\beta_1} \cdots \psi_g^{d\beta_g}.$$

This is a simple multiplicative equation,[2,3] and hence its solution is obtainable non-tentatively in a finite number of steps. The solution exhibits each of the u, ϕ, v, ψ as a power product in a certain minimum number of parameters θ, subject[2] to a set of G. C. D. conditions. The G. C. D. conditions on the ϕ, ψ occurring in the previous step are replaced by their equivalents in terms of the θ, on substituting for the ϕ, ψ their expressions as power products in the θ; if any θ occurs to a power higher than the first, this power is replaced by the first power.

A system of $\{\}$-equations leads in the same way to a simple multiplicative system, and hence the solution is obtainable non-tentatively in a finite number of steps.[2]

It remains to show how an equation of the form

$$\{z_1, \cdots, z_n\} = t_{n-1}, \qquad n > 1,$$

is to be solved for $z_1, \cdots, z_n, t_{n-1}$, and we shall assume first that the symbol $\{\}$ is simple. From this we shall obtain the solution when $\{\}$ is compound by iterating the method for $\{\}$ simple.

[3] M. Ward, ibid., 67–76.

4. Simple G. C. D. and L. C. M. equations

The abstract identity of $()'$, $[]'$ and $()$, $[]$ has been familiar for over a century, and Dedekind[4] (1897) referred to both as giving instances of his dual groups (structures, lattices), constructed on the operations now frequently called union and intersection; he also cited the classic instance, due to Boole, of logical sum and logical product. All three interpretations are required for a complete discussion of selective equations; the Boolean instance applies when we seek a minimum set of G. C. D. conditions, or their $()'$-equivalents, which imply the complete set as given by the method of solution. For the moment we are concerned only with $()$, $[]$; and we shall require the usual postulates

$$U(x, y) = U(y, x),$$
$$U(U(x, y), z) = U(x, U(y, z)),$$
$$U(x, C(x, y)) = x,$$

and the dual set obtained from these on replacing U by C and C by U. These are satisfied if $U(x, y)$ is the union of x, y and $C(x, y)$ their intersection; or if $U(x, y) \equiv (x, y)$, $C(x, y) \equiv [x, y]$; or if $U(x, y) \equiv (x, y)'$, $C(x, y) \equiv [x, y]'$.

The following recurrence relations are immediate consequences of the postulates,

$$(x_1, \cdots, x_n) = ((x_1, \cdots, x_{n-1}), x_n),\ n \geqq 2,$$
$$[y_1, \cdots, y_n] = [[y_1, \cdots, y_{n-1}], y_n],\ n \geqq 2;$$

and there is the well known identity $(x, y)[x, y] = xy$, which is evident from the duality in §2. It is required to solve the equations

$$(x_1, \cdots, x_n) = g_{n-1},\ n > 1;$$
$$[y_1, \cdots, y_n] = h_{n-1},\ n > 1.$$

From the recurrences and the identity we have the initial solutions ($n = 2$),

$$(x_1, x_2) = g_1:$$
$$x_1 = g_1u_1, \qquad x_2 = g_1u_2, \qquad 1 = (u_1, u_2);$$
$$[y_1, y_2] = h_1:$$
$$y_1 = k_1v_1, \qquad y_2 = k_1v_2, \qquad h_1 = k_1v_1v_2, \qquad 1 = (v_1, v_2).$$

From the first of these and the first recurrence we find

$$(x_1, \cdots, x_n) = g_{n-1},\ n > 2:$$
$$x_1 = g_{n-1}\prod_{i=1}^{n-1} u_{2i-1}, \qquad x_n = g_{n-1}u_{2n-2};$$

[4] R. Dedekind, Werke, 2, 112–114.

$$x_j = g_{n-1}u_{2j-2}\prod_{i=1}^{n-j} u_{2j+2i-3}, \qquad j = 2, \cdots, n-1;$$

$$1 = (u_1, u_2) = (u_3, u_4) = \cdots = (u_{2n-3}, u_{2n-2}),$$

in which the U's are integer parameters.

Similarly, from the second recurrence, we find

$$[y_1, \cdots, y_n] = h_{n-1}, n > 2:$$

$$y_1 = k_1v_1; \qquad y_t = k_{t-1}v_{2t-2}, t = 2, \cdots, n;$$

$$h_r = k_rv_{2r-1}v_{2r}, r = 1, \cdots, n-1;$$

$$1 = (v_1, v_2) = (v_3, v_4) = \cdots (v_{2n-3}, v_{2n-2}),$$

in which the integer parameters k, v are to be found from the complete solution of the simple multiplicative system[2]

$$k_sv_{2s-1}v_{2s} = k_{s+1}v_{2s+1}, s = 1, \cdots, n-2.$$

The typical equation has the complete solution

$$\begin{aligned} k_s &= af, & k_{s+1} &= abc,\\ v_{2s-1} &= bg, & v_{2s+1} &= fgh,\\ v_{2s} &= ch, & &\\ 1 &= (f, bc) = (b, fh). & & \end{aligned}$$

To solve an equation involving a compound symbol $\{\cdots, \{\}_1, \cdots, \}_n$, in which each of $\{\}_1, \cdots, \{\}_n$ is a definite one of (), [], we proceed from those $\{\}_j$ which enclose no symbol $\{\}$, say these are $\{\}_a, \cdots, \{\}_t$, and solve each of $\{\}_a = u, \cdots, \{\}_t = w$ by the above method. The $\{\}_a, \cdots, \{\}_t$ are then replaced in the original equation by the power products for $u, \cdots, w$ given by the solutions, and the process is repeated, until at the last step only a simple symbol $\{\}$ remains, when one more application of the solution for a simple equation gives the complete solution of the compound equation.

By §§2–4, the solution of the selective system in §1 is reduced to the solution of a simple multiplicative system, and this may be carried out non-tentatively in a finite number of steps.[2] After the detailed discussion, it is unnecessary to give examples (especially as those of greater interest have very long solutions), but it may be mentioned that those suggested by the Dedekind axiom, or by postulated distributivity of the operators considered, furnish interesting exercises.

An alternative method for dealing with $[y_1, \cdots, y_n]$, proposed by Lebesgue,[5] can be stated more briefly than that followed here, but actually it demands far more labor if $n > 4$. If $p_1 \equiv y_1 \cdots y_n$, and if p_j, $j > 1$, denotes the product of all the G. C. D.'s of the y's taken j at a time, Lebesgue's readily proved result is

$$[y_1, \cdots, y_n] = (p_1p_3p_5 \cdots)/(p_2p_4p_6 \cdots).$$

[5] V. A. Lebesgue, Nouv. Ann. Math., 8 (1849), 350.

A point of interest is that the G. C. D. conditions occurring in the solution of a simple multiplicative system by the method of arrays,[2] are *necessary* for the solution of a selective system. In the alternative method[3] for simple multiplicative systems, the G. C. D. conditions do not appear, and seem to have no relevance for the solution. As both methods furnish the complete solution of a simple multiplicative system, it is possible that an interpretation of selective systems in terms of the alternative method may suggest an appropriate equivalent of the G. C. D. conditions.

California Institute of Technology
Pasadena, California

ANNALS OF MATHEMATICS
Vol. 42, No. 4, October, 1941

SEMIGROUPS ADMITTING RELATIVE INVERSES

BY A. H. CLIFFORD

(Received December 17, 1940)

By a semigroup is meant a set S of elements $a, b, \cdots$ closed under an associative binary operation:

$$(ab)c = a(bc).$$

We shall say that S *admits relative inverses* if it satisfies the following condition.

I. *To each element a of S there exists an element e of S such that* (1) *e is an identity element of a*:

$$ea = ae = a,$$

and (2) *a possesses an inverse a' relative to e in S*:

$$aa' = a'a = e.$$

It is seen almost immediately (Theorem 1) that S admits relative inverses if and only if it is the class sum of mutually disjoint groups S_e, one to each idempotent element e of S. This fact tells us little about the structure of S since the product of two of these groups is not necessarily contained in a third, but may be scattered throughout several. (This occurs in the example given at the end of the paper, though simpler examples can be given.)

It is shown in Theorem 2, however, that S is the class sum of mutually disjoint semigroups S_α of known structure such that no such scattering takes place. Each S_α is what Rees[1] calls a completely simple semigroup without zero. The structure of such a semigroup was, in the finite case, first given by Suschkewitsch,[2] who calls it a *Kerngruppe*. Moreover it is possible to arrange these S_α in a semi-lattice[3] P such that the product $S_\alpha S_\beta$ is contained in the greatest lower bound (in P)$S_{\alpha\beta}$ of S_α and S_β. The structure of S is thus determined in the large, so to speak.

In the special case in which any two idempotent elements of S commute with each other, the simple semigroups S_α reduce to groups, coinciding with the groups S_e, and the semi-lattice P is isomorphic with the semigroup of all idempotent elements of S. In this case the structure of S is completely determined

[1] D. Rees, *On semi-groups* (Proc. Cambr. Phil. Soc. 36, 1940, 387–400).

[2] A. Suschkewitsch, *Über die endlichen Gruppen ohne das Gesetz der eindeutigen Umkehrbarkeit* (Math. Ann. 99, 1928, 30–50).

[3] A semi-lattice is a partially ordered set in which any two elements α, β have a greatest lower bound $\alpha\beta$, but not necessarily a least upper bound. F. Klein (Deutsche Math. 4, 1939, 32–43) calls it a *Halbverband*.

in Theorem 3. For every pair $\alpha > \beta$ there exists a homomorphism $\phi_{\alpha\beta}$ of the group S_α into the group S_β such that if $\alpha > \beta > \gamma$ then $\phi_{\alpha\gamma} = \phi_{\alpha\beta}\phi_{\beta\gamma}$. The product $a_\alpha b_\beta$ of any element a_α of S_α with any element b_β of S_β is then given by

$$a_\alpha b_\beta = (a_\alpha\phi_{\alpha\gamma})(b_\beta\phi_{\beta\gamma})$$

where $\gamma = \alpha\beta$. The groups S_α, the semi-lattice P, and the homomorphisms $\phi_{\alpha\beta}$ can (conversely) be chosen arbitrarily subject only to the above transitivity condition on the $\phi_{\alpha\beta}$.

In the general case, however, the situation is far more complicated. The concluding section discusses only the subsemigroup S' of S consisting of a pair S_α, S_β with $\alpha > \beta$. Theorem 4 cannot, even in this simple case, be regarded as a complete determination of S'; for it involves finding a certain mapping ϕ of S_α into the structure group of S_β the actual construction of which is not evident.

1. Decomposition into groups

LEMMA 1.1. *If a is any element of S, and e is any element of S satisfying* I, *then e is idempotent.*

PROOF: $e^2 = aa'aa' = aea' = aa' = e$.

LEMMA: 1.2. *If a is any element of S, and if e_1 and e_2 both satisfy* I, *then $e_1 = e_2$.*

PROOF: Let a_1 and a_2 be inverses of a relative to e_1 and e_2 respectively, as given by I(2):

$$aa_1 = a_1a = e_1, \qquad aa_2 = a_2a = e_2.$$

Then

$$e_1e_2 = e_1aa_2 = aa_2 = e_2,$$

$$e_1e_2 = a_1ae_2 = a_1a = e_1,$$

whence $e_1 = e_2$.

We shall say that a *belongs* to the uniquely determined idempotent element e satisfying I.

LEMMA 1.3. *The set S_e of all elements of S belonging to the idempotent element e of S is a group with identity e.*

PROOF: Let a and b be any two elements of S, and let a' and b' be inverses of a and b relative to e. Then e is evidently an identity element of ab, and $b'a'$ is an inverse of ab relative to e:

$$abb'a' = aea' = aa' = e,$$

$$b'a'ab = b'eb = b'b = e.$$

Hence $ab \in S_e$. Evidently $e \in S_e$, since $e^2 = e$, and e is an identity element of S_e. If a is any element of S_e, a' an inverse of a relative to e, and $b = ea'e$, then

$$eb = be = b,$$

$$ab = aea'e = aa'e = ee = e,$$

$$ba = ea'ea = ea'a = ee = e.$$

Hence $b \in S_e$ and b is an inverse of a therein. Thus S_e is a group.

Lemma 1.3 shows in particular that, among all possible inverses of a relative to the idempotent e to which a belongs, there is exactly one which also belongs to e. *This we shall denote by* a^{-1}.

Now by I every element of S belongs to at least one of the groups S_e, and by Lemma 1.2 to exactly one. Hence S is the class sum of the mutually disjoint groups S_e. We therefore have the following theorem,[4] the converse being evident.

THEOREM 1. *A semigroup admits relative inverses if and only if it is the class sum of mutually disjoint groups.*

2. Decomposition into completely simple semigroups

A subset $\mathfrak{a}$ of S is called an *ideal* if, for any a in $\mathfrak{a}$ and any x in S, ax and xa are in $\mathfrak{a}$; in other words, $\mathfrak{a}$ is an ideal if $S\mathfrak{a}S \subseteq \mathfrak{a}$. (We shall not have occasion to consider one-sided ideals.) The class sum $\mathfrak{a} \cup \mathfrak{b}$ and intersection $\mathfrak{a} \cap \mathfrak{b}$ of any two ideals $\mathfrak{a}$ and $\mathfrak{b}$ are evidently ideals, and likewise the set-product $\mathfrak{ab}$ consisting of all ab with a in $\mathfrak{a}$, b in $\mathfrak{b}$. Evidently

$$\mathfrak{ab} \subseteq \mathfrak{a} \cap \mathfrak{b}$$

and in particular $\mathfrak{a}^2 \subseteq \mathfrak{a}$. If a is any element of S, the set SaS of all elements xay of S is an ideal which we call the *principal ideal generated by* a. It contains a itself, since $a = eae$ by I, and hence contains aS and Sa.

LEMMA 2.1. *If an ideal* $\mathfrak{a}$ *contains a single element of the group* S_e *then it contains all of* S_e. *In particular, if* a^n *belongs to an ideal* $\mathfrak{a}$, *then* a *itself belongs to* $\mathfrak{a}$.

PROOF: Let a be an element in both $\mathfrak{a}$ and S_e. Then $\mathfrak{a}$ contains $a^{-1}a = e$. If b is any other element of S_e, then $\mathfrak{a}$ contains $be = b$. Hence $\mathfrak{a} \supseteq S_e$. The second statement then follows from the fact that a^n belongs to the same group S_e as a.

LEMMA 2.2. *If* $\mathfrak{a}$ *and* $\mathfrak{b}$ *are ideals of* S *then*

$$\mathfrak{ab} = \mathfrak{a} \cap \mathfrak{b}.$$

In particular, $\mathfrak{ab} = \mathfrak{ba}$ *and every ideal* $\mathfrak{a}$ *of* S *is idempotent:* $\mathfrak{a}^2 = \mathfrak{a}$.

PROOF: We prove the last statement first. Let a be any element of $\mathfrak{a}$. Then a^2 is in $\mathfrak{a}^2$, whence a is in $\mathfrak{a}^2$ by Lemma 2.1. Thus $\mathfrak{a} \subseteq \mathfrak{a}^2$ whence $\mathfrak{a} = \mathfrak{a}^2$. Hence if $\mathfrak{a}$ and $\mathfrak{b}$ are any two ideals of S,

$$\mathfrak{a} \cap \mathfrak{b} = (\mathfrak{a} \cap \mathfrak{b})(\mathfrak{a} \cap \mathfrak{b}) \subseteq \mathfrak{a}\cdot\mathfrak{b}$$

from which equality follows.

LEMMA 2.3. *The product of two principal ideals* SaS *and* SbS *of* S *is the principal ideal* $SabS$. *If* a *and* b *generate the same principal ideal* $\mathfrak{a}$, *then* ab *is also a generator of* $\mathfrak{a}$.

PROOF: If x is any element of S,

$$(bxa)^2 = bx\cdot ab\cdot xa \in SabS.$$

[4] This result generalizes Theorem 14 of A. R. Poole, *Finite Ova* (Amer. Jour. of Math. 59, 1937, 23–32).

By Lemma 2.1, $bxa \in SabS$. Hence

$$bSa \subseteq SabS$$

and by Lemma 2.2,

$$SaS \cdot SbS = SbS \cdot SaS \subseteq S \cdot bSa \cdot S \subseteq SabS.$$

The opposite inclusion is evident.

If $SaS = SbS = \mathfrak{a}$, then from the above and Lemma 2.2,

$$SabS = SaS \cdot SbS = \mathfrak{a}^2 = \mathfrak{a}.$$

LEMMA 2.4. *If $\mathfrak{a}$ is a principal ideal of S then the set $\mathfrak{a}'$ of all elements of $\mathfrak{a}$ which are not generators of $\mathfrak{a}$ is also an ideal of S.*

PROOF: $S\mathfrak{a}'S \subseteq SaS \subseteq \mathfrak{a}$; we wish to show that $S\mathfrak{a}'S \subseteq \mathfrak{a}'$. Suppose $S\mathfrak{a}'S$ contained an element a of $\mathfrak{a}$ not in $\mathfrak{a}'$. Then $a = xa'y$ with a' in $\mathfrak{a}'$, and a generates $\mathfrak{a}$. Hence

$$\mathfrak{a} = SaS = Sxa'yS \subseteq Sa'S;$$

thus a' would be a generator of $\mathfrak{a}$, contrary to $a' \in \mathfrak{a}'$.

An element 0 of a semigroup S is called a *zero element* of S if $0a = a0 = 0$ for all a in S. Rees (loc. cit. p. 392) defines a simple semigroup to be a semigroup S whose only ideals are S and the null ideal (0) consisting of 0 alone (if S has a zero). By a *simple semigroup without zero* we shall mean a semigroup S whose only ideal is S itself, including thereby the case in which S consists of a single element. Such a semigroup is characterized by the property that $SaS = S$ for any element a of S, or that, for given a and b in S, the equation $xay = b$ is always solvable for x and y in S.

LEMMA 2.5. *If $\mathfrak{a}$ is a principal ideal of S, then the set $S_\mathfrak{a}$ of all generators of $\mathfrak{a}$ is a simple semigroup without zero and admits relative inverses.*

PROOF: $S_\mathfrak{a}$ is a subsemigroup of S by Lemma 2.3. Let a and b be arbitrary elements of $S_\mathfrak{a}$; we are to show that $xay = b$ is solvable for x and y in $S_\mathfrak{a}$.

Now $\mathfrak{a}a\mathfrak{a}$ is an ideal containing a^3, and hence containing a by Lemma 2.1. Since a generates $\mathfrak{a}$ and $\mathfrak{a}a\mathfrak{a} \subseteq \mathfrak{a}$, we have $\mathfrak{a}a\mathfrak{a} = \mathfrak{a}$. Hence $b = xay$ with x and y in $\mathfrak{a}$. By Lemma 2.4, x and y must both lie in $S_\mathfrak{a}$; for if either of them lay in the ideal $\mathfrak{a}' = \mathfrak{a} - S_\mathfrak{a}$ then the same would be true of b, contrary to $b \in S_\mathfrak{a}$. Hence $S_\mathfrak{a}$ is a simple semigroup without zero.

If a is any element of $S_\mathfrak{a}$, and e is the idempotent to which a belongs, then by Lemma 2.1, $S_e \subseteq \mathfrak{a}$. By the same lemma, no element of S_e can belong to the ideal $\mathfrak{a}'$. Hence $S_e \subseteq S_\mathfrak{a}$. $S_\mathfrak{a}$ is thus the class sum of certain of the groups S_e, and so admits relative inverses.

If e and f are idempotents of S, then e is said to be *under* f (Rees p. 393) if $ef = fe = e$. We shall write $e \leqq f$ if e is under f; the relation $\leqq$ is easily seen to be a partial ordering of the set of idempotents of S. Rees makes the following definitions (p. 393).

An idempotent f is *primitive* if there exists no idempotent $e \neq 0$ such that $e < f$. A semigroup S is *completely simple* if

(i) *S is simple;*

(ii) *to each a in S there exist idempotents e and f in S such that $ea = af = a$;*

(iii)′ *every idempotent of S is primitive.* (We take here his condition (iii)′ on p. 394 which he proves to be equivalent to (iii).)

Rees (p. 399) shows that any completely simple semigroup is isomorphic with what he calls (p. 397) a *regular matrix semigroup over a group with zero.* When no zero element is present, the latter may be described as follows.

Let G be any group. Let J and K be any two sets of indices; denote the elements of J by $i, j, \cdots$, and those of K by $\kappa, \lambda, \cdots$. To each pair i, κ of indices ($i \in J, \kappa \in K$) assign an arbitrary but fixed element $p_{\kappa i}$ of G. The semigroup S shall consist of all triples ("matrices")

$$(a; i, \kappa) \qquad (a \in G; i \in J, \kappa \in K)$$

with multiplication defined by

$$(a; i, \kappa)(b; j, \lambda) = (ap_{\kappa j}b; i, \lambda). \tag{2.1}$$

The idempotents of S are readily seen from (2.1) to be the elements $(p_{\kappa i}^{-1}; i, \kappa)$.

LEMMA 2.6. *If S is a semigroup admitting relative inverses, and if e and f are idempotents of S such that $e \leqq f$ and f lies in SeS, then $e = f$.*

PROOF: From $f \in SeS$ we have $f = xey$ for some x and y in S. Setting $a = fxf$, $b = fyf$, and using $fe = ef = e$, we have

$$aeb = fxfefyf = fxeyf = fff = f$$

together with

$$fa = af = a, \qquad fb = bf = b.$$

Let g be the idempotent to which a belongs. Then

$$f = aeb = gaeb = gf = a^{-1}af = a^{-1}a = g.$$

Hence a belongs to f, and similarly b belongs to f. Consequently

$$a^{-1}fb^{-1} = a^{-1}aebb^{-1} = fef = e.$$

Since a^{-1} and b^{-1} are in S_f, this implies that e is in S_f, whence $e = f$.

LEMMA 2.7. *A simple semigroup S without zero is completely simple if and only if it admits relative inverses.*

PROOF: Assume that S admits relative inverses. Condition (i) for complete simplicity holds by hypothesis. (ii) follows from I(1), and (iii)′ from Lemma 2.6, since $SeS = S$.

Conversely, if S is completely simple then it has the matrix structure described above. If $(a; i, \kappa)$ is any element of S one readily verifies by direct calculation from (2.1) that $(p_{\kappa i}^{-1}; i, \kappa)$ is an identity and $(p_{\kappa i}^{-1}a^{-1}p_{\kappa i}^{-1}; i, \kappa)$ a relative inverse of $(a; i, \kappa)$.

By a *semi-lattice* (Klein, loc. cit.) we shall mean a commutative semigroup P

all of whose elements are idempotent. The reason for this term is that if (as above) we define $\alpha \leqq \beta$ (α and β in P) to mean $\alpha\beta = \alpha$, then P is a partially ordered set under $\leqq$ in which every pair of elements α and β have a greatest lower bound, namely $\alpha\beta$. The converse is likewise evident.

Now Lemmas 2.2 and 2.3 show that the set $\mathfrak{P}$ of principal ideals of S is a semi-lattice (actually $\mathfrak{a}\mathfrak{b} = \mathfrak{a} \cap \mathfrak{b}$ and $\mathfrak{a} \leqq \mathfrak{b}$ is equivalent to $\mathfrak{a} \subseteq \mathfrak{b}$). To each $\mathfrak{a}$ in $\mathfrak{P}$ corresponds the semigroup $S_\mathfrak{a}$ of all generators of $\mathfrak{a}$. If $a \in S_\mathfrak{a}$, $b \in S_\mathfrak{b}$ then by Lemma 2.3, ab generates $\mathfrak{a}\mathfrak{b}$ and so belongs to $S_{\mathfrak{a}\mathfrak{b}}$. Hence $S_\mathfrak{a} S_\mathfrak{b} \subseteq S_{\mathfrak{a}\mathfrak{b}}$. Every element a of S is a generator of exactly one principal ideal $\mathfrak{a} = SaS$ of S, and so belongs to exactly one $S_\mathfrak{a}$. Using Lemmas 2.5 and 2.7 we arrive at the following description of the gross structure of S.

THEOREM 2. *Every semigroup S admitting relative inverses determines a semi-lattice P such that to each element α of P there corresponds a subsemigroup S_α of S with the following properties.*

(1) *The S_α are mutually disjoint and their class sum is S.*

(2) *Each S_α is a completely simple semigroup without zero, hence a matrix semigroup over a group.*

(3) *$S_\alpha S_\beta \subseteq S_{\alpha\beta}$, where $\alpha\beta$ is the product of α and β in P.*

Conversely, any semigroup having this structure admits relative inverses.

We do not mention here the isomorphism of P with the semi-lattice of principal ideals of S. But if an ideal of S contains a single element of S_α it must by (2) contain the whole of S_α. From this it is clear that if S is any semigroup having the structure described in Theorem 2, then the lattice of all its ideals is isomorphic with the lattice of all ideals of P.

3. Case in which the idempotents of S commute

LEMMA 3.1. *If S is a semigroup admitting relative inverses, and if every pair of idempotent elements of S commute with each other, then every idempotent of S is in the center of S.*

PROOF: The set $\mathfrak{E}$ of idempotents of S evidently forms a commutative sub-semigroup of S, hence a semi-lattice as defined above. We prove the lemma in four stages; e and f denote idempotents, as usual.

(1) If $SeS = SfS$ then $e = f$. For $ef \leqq e$ and $e \in SefS$ since $SefS = SeS$ by Lemma 2.3. Hence $ef = e$ by Lemma 2.6, that is, $e \leqq f$. By symmetry, $f \leqq e$, whence $e = f$.

(2) If $SaS = SeS$ then $a \in S_e$. For if $a \in S_f$ then $SaS = SfS$ by Lemma 2.1, whence $f = e$ by (1).

(3) If $a \in S_f$ and $e \leqq f$, then $ae = ea$. For $e = fe = a^{-1}ae$ and so $SeS \subseteq SaeS$. Since the opposite inclusion is evident, $SaeS = SeS$. Similarly $SeaS = SeS$. Hence ae and ea belong to S_e by (2), whence $ae = e \cdot ae = ea \cdot e = ea$.

(4) Now let e and a be arbitrary. Suppose a belongs to S_f. Since $ef \leqq f$ it follows from (3) that $a \cdot ef = ef \cdot a$. But

$$a \cdot ef = a \cdot fe = af \cdot e = ae,$$

$$ef \cdot a = e \cdot fa = ea,$$

whence $ae = ea$.

Part (2) in the above proof shows that every generator of $\mathfrak{a} = SaS$ lies in S_e, and hence $S_{\mathfrak{a}} = S_e$. Thus the simple semigroups $S_{\mathfrak{a}}$ reduce to the groups S_e in this case, and the semi-lattice $\mathfrak{P}$ is isomorphic with $\mathfrak{E}$. We shall, as in Theorem 2, denote the elements of the abstract semi-lattice P isomorphic with $\mathfrak{P}$ and $\mathfrak{E}$ by $\alpha, \beta, \gamma, \cdots$, and the elements of $\mathfrak{E}$ correspondingly by $e_\alpha, e_\beta, \cdots$. We thus have

$$e_\alpha e_\beta = e_{\alpha\beta} = e_{\beta\alpha} = e_\beta e_\alpha$$

and $e_\alpha \geqq e_\beta$ if and only if $\alpha \geqq \beta$. S_α is the same as S_{e_α} and

$$S_\alpha S_\beta \subseteq S_{\alpha\beta}.$$

In particular, if $\alpha \geqq \beta$ then $S_\alpha S_\beta \subseteq S_\beta$ and $e_\alpha e_\beta = e_\beta$. Elements of S_α will be denoted by $a_\alpha, b_\alpha, \cdots$.

LEMMA 3.2. *If $\alpha \geqq \beta$, the mapping $a_\alpha \to a_\alpha \phi_{\alpha\beta}$ defined by*

$$a_\alpha \phi_{\alpha\beta} = a_\alpha e_\beta$$

is a homomorphism of S_α into S_β. If $\alpha \geqq \beta \geqq \gamma$ then the homomorphism $\phi_{\alpha\gamma}$ of S_α into S_γ is the product of the homomorphisms $\phi_{\alpha\beta}$ and $\phi_{\beta\gamma}$ of S_α into S_β and S_β into S_γ:

$$\phi_{\alpha\gamma} = \phi_{\alpha\beta}\phi_{\beta\gamma} \qquad (\alpha \geqq \beta \geqq \gamma). \tag{3.1}$$

$\phi_{\alpha\alpha}$ is the identical automorphism of S_α.

PROOF: That $\phi_{\alpha\beta}$ maps S_α into S_β is clear. If a_α and b_α are arbitrary elements of S_α then (using Lemma 3.1)

$$(a_\alpha\phi_{\alpha\beta})(b_\alpha\phi_{\alpha\beta}) = a_\alpha e_\beta b_\alpha e_\beta = a_\alpha b_\alpha e_\beta = (a_\alpha b_\alpha)\phi_{\alpha\beta}.$$

If $\alpha \geqq \beta \geqq \gamma$ then for any a_α in S_α,

$$(a_\alpha\phi_{\alpha\beta})\phi_{\beta\gamma} = (a_\alpha e_\beta)e_\gamma = a_\alpha(e_\beta e_\gamma) = a_\alpha e_\gamma = a_\alpha\phi_{\alpha\gamma}.$$

Finally, $a_\alpha\phi_{\alpha\alpha} = a_\alpha e_\alpha = a_\alpha$, for any a_α in S_α.

LEMMA 3.3. *If a_α and b_β are any two elements of S (a_α in S_α, b_β in S_β) then*

$$a_\alpha b_\beta = (a_\alpha\phi_{\alpha\gamma})(b_\beta\phi_{\beta\gamma}) \tag{3.2}$$

where $\gamma = \alpha\beta$.

PROOF:
$$a_\alpha b_\beta = a_\alpha e_\alpha \cdot b_\beta e_\beta = a_\alpha b_\beta e_\alpha e_\beta = a_\alpha b_\beta e_\gamma$$
$$= a_\alpha e_\gamma \cdot b_\beta e_\gamma = (a_\alpha\phi_{\alpha\gamma})(b_\beta\phi_{\beta\gamma}).$$

From these two lemmas it is clear that the structure of S is completely known when we know the semi-lattice P, the groups S_α, and the homomorphisms $\phi_{\alpha\beta}$.

Conversely, let P be any semi-lattice, and to each α in P assign a group S_α such that no two of them have an element in common. Furthermore to each pair of elements α, β of P such that $\alpha > \beta$ assign a homomorphism $\phi_{\alpha\beta}$ of S_α into S_β. Define $\phi_{\alpha\alpha}$ to be the identical automorphism of S_α. Let S be the class sum of the groups S_α. Then, if the homomorphisms $\phi_{\alpha\beta}$ satisfy the transitivity conditions (3.1) and we define multiplication in S by (3.2), S is an associative semigroup.

For if a_α, b_β, c_γ are any three elements of S,

$$(a_\alpha b_\beta)c_\gamma = (a_\alpha\phi_{\alpha,\alpha\beta}\cdot b_\beta\phi_{\beta,\alpha\beta})c_\gamma$$

$$= (a_\alpha\phi_{\alpha,\alpha\beta}\cdot b_\beta\phi_{\beta,\alpha\beta})\phi_{\alpha\beta,\alpha\beta\gamma}\cdot c_\gamma\phi_{\gamma,\alpha\beta\gamma}$$

since the first factor on the right is in $S_{\alpha\beta}$. Using first the fact that $\phi_{\alpha\beta,\alpha\beta\gamma}$ is a homomorphism and then (3.1), we get

$$(a_\alpha b_\beta)c_\gamma = (a_\alpha\phi_{\alpha,\alpha\beta\gamma}\cdot b_\beta\phi_{\beta,\alpha\beta\gamma})\cdot c_\gamma\phi_{\gamma,\alpha\beta\gamma}.$$

Similarly,

$$a_\alpha(b_\beta c_\gamma) = a_\alpha\phi_{\alpha,\alpha\beta\gamma}\cdot(b_\beta\phi_{\beta,\alpha\beta\gamma}\cdot c_\gamma\phi_{\gamma,\alpha\beta\gamma}),$$

and these are equal by the associative law in $S_{\alpha\beta\gamma}$.

The semigroup S so constructed evidently admits relative inverses, since it is a sum of groups. Likewise it is evident that the only idempotent elements of S are the identity elements e_α of the groups S_α. Since $\phi_{\alpha\beta}$ is a homomorphism, $e_\alpha\phi_{\alpha\beta}$ must be the identity element e_β of S_β. Hence if e_α is any idempotent element of S, and b_β any element of S,

$$\begin{aligned} e_\alpha b_\beta &= (e_\alpha\phi_{\alpha\gamma})(b_\beta\phi_{\beta\gamma}) = e_\gamma(b_\beta\phi_{\beta\gamma}) = b_\beta\phi_{\beta\gamma}, \\ b_\beta e_\alpha &= (b_\beta\phi_{\beta\gamma})(e_\alpha\phi_{\alpha\gamma}) = (b_\beta\phi_{\beta\gamma})e_\gamma = b_\beta\phi_{\beta\gamma}. \end{aligned} \qquad (\gamma = \alpha\beta)$$

Thus e_α is in the center of S.

Finally we remark that the semi-lattice P, the groups S_α, and the homomorphisms $\phi_{\alpha\beta}$ constitute a complete set of invariants of S. To see that they are invariants, suppose $S \cong \bar{S}$. Since idempotents must correspond to idempotents, $\mathfrak{E} \cong \bar{\mathfrak{E}}$, and the same abstract P may serve for both. But if $e \leftrightarrow \bar{e}$ then S_e and $\bar{S}_{\bar{e}}$ correspond, so that S_α and $\bar{S}_\alpha$ are isomorphic. Since

$$\bar{a}_\alpha\bar{\phi}_{\alpha\beta} = \bar{a}_\alpha\bar{e}_\beta \leftrightarrow a_\alpha e_\beta = a_\alpha\phi_{\alpha\beta}$$

the homomorphisms $\bar{\phi}$ are cogredient with the homomorphisms ϕ in the sense that if $a \leftrightarrow \bar{a}$ then $a\phi \leftrightarrow \bar{a}\bar{\phi}$. On the other hand, if S and $\bar{S}$ have the same invariants in the above sense, we may identify $\bar{P}$ with P and then by Lemma 3.3 the given isomorphisms between the groups S_α and $\bar{S}_\alpha$ define an isomorphism between S and $\bar{S}$.

THEOREM 3. *Every semigroup which admits relative inverses and in which every pair of idempotent elements commute is isomorphic with a semigroup S constructed as follows.*

Let P be any semi-lattice, and to each α in P assign a group S_α such that no two of them have an element in common. To each pair of elements $\alpha > \beta$ of P assign a homomorphism $\phi_{\alpha\beta}$ of S_α into S_β such that if $\alpha > \beta > \gamma$ then

$$\phi_{\alpha\beta}\phi_{\beta\gamma} = \phi_{\alpha\gamma}.$$

Let $\phi_{\alpha\alpha}$ be the identical automorphism of S_α. Let S be the class sum of the groups S_α, and define the product of any two elements a_α, b_β of S (a_α in S_α and b_β in S_β) by

$$a_\alpha b_\beta = (a_\alpha \phi_{\alpha\gamma})(b_\beta \phi_{\beta\gamma})$$

where $\gamma = \alpha\beta$ is the product of α and β in P.

Conversely, any semigroup S constructed in this fashion admits relative inverses, and every idempotent element of S is in the center of S.

The semi-lattice P, the groups S_α, and the homomorphisms $\phi_{\alpha\beta}$ together constitute a complete set of invariants of S.

4. Structure of a pair S_α, S_β with $\alpha > \beta$

We return now to the general case described in Theorem 2, and consider the structure of the subsemigroup S' of S consisting of a pair S_α, S_β with $\alpha > \beta$. For this purpose we may evidently assume that $S = S'$. As far as the results of this section go, S_α may be any semigroup, while S_β is to have the structure described in §2, the elements of S_β being the triples $(a; i, \kappa)$ with multiplication defined by (2.1). We shall denote the elements of S_α by capital letters $A, B, \cdots$.

Suppose that

$$A(a; i, \kappa) = (a'; i', \kappa').$$

Multiplying on the right by the idempotent $(p_{\kappa i}^{-1}; i, \kappa)$ to which $(a; i, \kappa)$ belongs, we get

$$A(a; i, \kappa) = (a' p_{\kappa' i} p_{\kappa i}^{-1}; i', \kappa),$$

whence $\kappa = \kappa'$. Hence

$$A(a; i, \kappa) = (a'; i', \kappa)$$

and similarly

$$(a; i, \kappa)A = (a''; i, \kappa'').$$

Suppose now that

$$A(p_{\kappa i}^{-1}; i, \kappa) = (a'; i', \kappa),$$

$$A(p_{\lambda i}^{-1}; i, \lambda) = (a''; i'', \lambda).$$

Multiplying the first of these on the right by $(p_{\lambda i}^{-1}; i, \lambda)$ we obtain

$$A(p_{\lambda i}^{-1}; i, \lambda) = (a' p_{\kappa i} p_{\lambda i}^{-1}; i', \lambda).$$

Hence

$$i'' = i', \qquad a'' = a' p_{\kappa i} p_{\lambda i}^{-1}.$$

The first of these shows that i' depends only on A and i and not on κ; we may

therefore write $i' = Ai$. The second shows that $a'p_{\kappa i}$ is likewise independent of κ; we may therefore write

$$a'p_{\kappa i} = A\phi_i$$

where ϕ_i is some mapping (depending on i) of S_α into the structure group G of S_β. We then have

$$A(p_{\kappa i}^{-1}\,; i, \kappa) = (A\phi_i \cdot p_{\kappa i}^{-1}\,; Ai\,, \kappa).$$

Finally, multiplying on the right by $(a; i, \kappa)$,

$$A(a; i, \kappa) = (A\phi_i \cdot a; Ai, \kappa). \tag{4.1}$$

By left-right duality, there likewise exist mappings

$$\kappa \rightarrow \kappa A \text{ of } K \text{ into itself,}$$

$$A \rightarrow A\psi_\kappa \text{ of } S_\alpha \text{ into } G,$$

such that

$$(a; i, \kappa)A = (a \cdot A\psi_\kappa\,; i, \kappa A). \tag{4.2}$$

By double applications of (4.1) and (4.2) we find

$$AB(a; i, \kappa) = (A\phi_{Bi} \cdot B\phi_i \cdot a; A(Bi), \kappa),$$

$$(a; i, \kappa)AB = (a \cdot A\psi_\kappa \cdot B\psi_{\kappa A}\,; i, (\kappa A)B).$$

Hence

$$(AB)i = A(Bi), \qquad \kappa(AB) = (\kappa A)B; \tag{4.3}$$

$$(AB)\phi_i = (A\phi_{Bi})(B\phi_i), \qquad (AB)\psi_\kappa = (A\psi_\kappa)(B\psi_{\kappa A}). \tag{4.4}$$

Likewise

$$(a; j, \kappa)A \cdot (b; i, \lambda) = (a \cdot A\psi_\kappa \cdot p_{\kappa A, i} b; j, \lambda),$$

$$(a; j, \kappa) \cdot A(b; i, \lambda) = (ap_{\kappa, Ai} \cdot A\phi_i \cdot b; j, \lambda),$$

whence

$$(A\psi_\kappa)p_{\kappa A, i} = p_{\kappa, Ai}(A\phi_i). \tag{4.5}$$

Conversely, let S_α be any semigroup and S_β a matrix semigroup over a group G as above. Let there be given a left representation of S_α by mappings of J into itself, and a right representation of S_α by mappings of K into itself, in other words (4.3). To each i in J (and each κ in K) let there be given a mapping $\phi_i(\psi_\kappa)$ of S_α into G such that (4.4) and (4.5) hold. Let S be the class sum of S_α and S_β. Then if we define multiplication between elements of S_α and S_β by (4.1) and (4.2), the associative law holds, and S is a semigroup. Of the six possible cases to consider, three are evident from the derivation of (4.3), (4.4), (4.5), and the remaining three are identically satisfied, as is easily seen.

The equations (4.4) and (4.5) can be simplified by normalizing[5] the $p_{\kappa i}$-matrix so that

$$p_{\kappa 1} = p_{1i} = e$$

where e is the identity element of G. (No confusion arises by denoting these particular elements of J and K by the same letter 1.) Rather than refer to Rees's Theorem 2.91 and the notion of equivalence, simply observe that if we write

$$[a; i, \kappa] = (q_i a r_\kappa\, ; i, \kappa)$$

then

$$[a; j, \kappa][b; i, \lambda] = [a r_\kappa p_{\kappa i} q_i b; j, \lambda].$$

Thus $p_{\kappa i}$ is replaced by

$$p'_{\kappa i} = r_\kappa p_{\kappa i} q_i\, .$$

If we choose

$$r_\kappa = p_{11} p_{\kappa 1}^{-1}\,, \qquad q_i = p_{1i}^{-1}$$

then $p'_{1i} = p'_{\kappa 1} = e$.

Assuming this normalization and setting $\kappa = 1$, $i = 1$ in (4.5) we obtain $A\psi_1 = A\phi_1$. Denote $\psi_1 = \phi_1$ by ϕ. Setting $\kappa = 1$ and $i = 1$ separately in (4.5) we find

$$(4.6) \qquad A\phi_i = (A\phi)p_{1A,i}\,, \qquad A\psi_\kappa = p_{\kappa,A1}(A\phi).$$

Putting these back in (4.5),

$$(4.7) \qquad p_{\kappa,A1}(A\phi)p_{\kappa A,i} = p_{\kappa,Ai}(A\phi)p_{1A,i}\,.$$

Setting $i = 1$ (or $\kappa = 1$) in (4.4) and using (4.6),

$$(4.8) \qquad (AB)\phi = (A\phi)p_{1A,B1}(B\phi).$$

(4.1) and (4.2) become respectively

$$(4.9) \qquad A(a; i, \kappa) = (A\phi\cdot p_{1A,i}a; Ai, \kappa),$$

$$(4.10) \qquad (a; i, \kappa)A = (ap_{\kappa,A1}\cdot A\phi; i, \kappa A).$$

THEOREM 4. *Let S be a semigroup which splits into two disjoint semigroups S_α and S_β such that $S_\alpha S_\beta$ and $S_\beta S_\alpha$ are contained in S_β, and where S_β is a completely simple semigroup without zero. If in the representation of S_β as a matrix semigroup over a group G we normalize the defining matrix $(p_{\kappa i})$ so that $p_{\kappa 1} = p_{1i} = e$, the identity element of G, then there exist left and right representations* (4.3) *of S_α by*

[5] That this normalization is possible was shown by Suschkewitsch (loc. cit.) in the finite case.

mappings of the index classes J and K of S_β into themselves, and a mapping ϕ of S_α into G satisfying (4.7) *and* (4.8), *such that the products in $S_\alpha S_\beta$ and $S_\beta S_\alpha$ are given by* (4.9) *and* (4.10).

Conversely, let S_α be any semigroup and S_β a completely simple semigroup without zero having no element in common with S_α. Let S_β be given as a matrix semigroup over a group G with matrix $(p_{\kappa i})$ normalized as above. Let there be given left and right representations (4.3) *of S_α by mappings of the index classes J and K of S_β into themselves, and a mapping ϕ of S_α into G satisfying* (4.7) *and* (4.8). *Let S be the class sum of S_α and S_β. Then* (4.9) *and* (4.10) *define an associative multiplication in S such that $S_\alpha S_\beta$ and $S_\beta S_\alpha$ are contained in S_β.*

PROOF: The first part has of course already been shown. To show the second part, define $A\phi_i$ and $A\psi_\kappa$ by (4.6), so that (4.9) and (4.10) go back into (4.1) and (4.2). We already have (4.3) and so need prove only (4.4) and (4.5) for associativity. Using the definition (4.6) we have

$$\begin{aligned}(AB)\phi_i &= (AB)\phi\cdot p_{1AB,i}\\ &= (A\phi)p_{1A,B1}(B\phi)p_{1AB,i}\end{aligned}$$

by (4.8). In (4.7) replace A by B and κ by $1A$:

$$p_{1A,B1}(B\phi)p_{1AB,i} = p_{1A,Bi}(B\phi)p_{1B,i}.$$

Hence

$$\begin{aligned}(AB)\phi_i &= (A\phi)p_{1A,Bi}(B\phi)p_{1B,i}\\ &= (A\phi_{Bi})(B\phi_i).\end{aligned}$$

The second part of (4.4) is shown in a similar manner. To show (4.5) we have by (4.7)

$$\begin{aligned}(A\psi_\kappa)p_{\kappa A,i} &= p_{\kappa,A1}(A\phi)p_{\kappa A,i},\\ &= p_{\kappa,Ai}(A\phi)p_{1A,i}\\ &= p_{\kappa,Ai}(A\phi_i).\end{aligned}$$

That $S_\alpha S_\beta$ and $S_\beta S_\alpha$ are contained in S_β is obvious.

If all the $p_{\kappa i} = e$, which is the case if and only if the product of any two idempotents of S_β is idempotent, (4.7) is no condition at all on ϕ, and (4.8) reduces to

$$(AB)\phi = (A\phi)(B\phi).$$

Hence the structure of S is determined in this case by the representations (4.3) and a homomorphism ϕ of S_α into G.

We close with an example to show that in the contrary case ϕ need not be a homomorphism. Let G be a cyclic group of order three, $G = \{e, a, a^2\}$, $a^3 = e$. Let $J = K = \{1, 2\}$. Let $p_{11} = p_{12} = p_{21} = e$, $p_{22} = a$. Let S_α be a cyclic group of order two, $S_\alpha = \{E, A\}$, $A^2 = E$. Let

$$E1 = 1E = 1, \qquad E2 = 2E = 2,$$
$$A1 = 1A = 2, \qquad A2 = 2A = 1.$$

Finally let

$$E\phi = e, \qquad A\phi = a.$$

This is evidently not a homomorphism of S_α into G. Yet the conditions (4.7) and (4.8) are easily verified. For example,

$$(AA)\phi = E\phi = e,$$

$$(A\phi)p_{1A,A1}(A\phi) = ap_{22}a = a^3 = e.$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

The publication of the supplementary number
is made possible by a grant from the
American Philosophical Society

Annals of Mathematics
Vol. 42, No. 5, December, 1941

DIE TOPOLOGIE DER LIESCHEN GRUPPEN ALS ALGEBRAISCHES PHÄNOMEN. I

Von Hans Freudenthal

(Received November 15, 1940)

Daß man die Topologie der Lieschen Gruppen nicht rein gruppentheoretisch erfassen kann, ist bekannt. Desto merkwürdiger ist der zuerst von E. Cartan,[1] danach (ohne Kenntnis der Cartanschen Arbeit) andersartig von B. L. v. d. Waerden[2] bewiesene Satz:

Ein Isomorphismus zwischen zwei *halbeinfachen kompakten* Lieschen Gruppen ist notwendig stetig.

Daß die Halbeinfachheit allein die Gültigkeit des Satzes nicht gewährleistet, zeigt das laut v. d. Waerden von J. v. Neumann herrührende Beispiel:

In der Gruppe $\mathfrak{g}$ der komplexen Matrices $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ mit der Determinante 1 kann man durch unstetige Automorphismen f des Körpers der komplexen Zahlen unstetige Automorphismen

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \to \begin{pmatrix} f(\alpha) & f(\beta) \\ f(\gamma) & f(\delta) \end{pmatrix}$$

definieren.

Diese Gruppe $\mathfrak{g}$ ist nun als Liesche Gruppe von 6 reellen Parametern ($\mathfrak{R}_\alpha$, $\mathfrak{R}_\beta$, $\mathfrak{R}_\gamma$, $\mathfrak{J}_\alpha$, $\mathfrak{J}_\beta$, $\mathfrak{J}_\gamma$) einfach (bis auf einen diskreten Normalteiler), und doch fehlt dem Beispiel die rechte Überzeugungskraft. Und das kommt so: Bei der reellen Klassifikation der *reellen einfachen* nichtabelschen Gruppen[3] unterscheidet man die

erster Art, die *komplex nicht mehr einfach* (sondern direktes Produkt zweier konjugierter einfacher Gruppen sind), und die

zweiter Art, die *auch komplex einfach* sind.

Die erster Art lassen sich, wie man leicht beweist,[4] alle so erzeugen: Man nehme eine reelle einfache nichtabelsche Gruppe, setze sie ins Komplexe fort und trenne Real- und Imaginärteil; diesen Prozeß, der die Dimensionszahl verdoppelt, wollen wir "Verdoppelung" nennen.

Eine Gruppe erster Art haben wir gerade in obigem Beispiel angetroffen. So

[1] E. Cartan, Sur les représentations linéaires des groupes clos [Comment. Helvet. **2** (1930), 269–283].

[2] B. L. van der Waerden, Stetigkeitssätze für halbeinfache Liesche Gruppen [Math. Zeitschr. **36** (1933), 780–786].

[3] E. Cartan, Les groupes réels simples, finis et continus [Ann. Ecole Norm. Sup. (3) **31** (1914), 263–355].

[4] a. a. O.,[3] 266–267.

beschleicht uns der Verdacht, daß für den Cartan-v.d.Waerdenschen Satz nicht die Kompaktheit, sondern die Zugehörigkeit der einfachen Gruppe zur zweiten Art verantwortlich sei.

In der Tat beweisen wir den

HAUPTSATZ I: *Ein Isomorphismus zwischen zwei Lieschen Gruppen, von denen die eine einfach und von zweiter Art ist, ist notwendig stetig. Oder: Im Bereiche der Lieschen Gruppen ist die Topologie der einfachen Gruppen zweiter Art ein rein algebraisches Phänomen.*

(In der letzten Formulierung ist der Hauptsatz auch sinnvoll für jemanden, der unstetige Abbildungen überhaupt ablehnt—etwa aus intuitionistischen Erwägungen.)

BEMERKUNG 1: Der Hauptsatz bleibt gültig, wenn man die Einfachheit nur im Kleinen fordert.

BEMERKUNG 2: Hauptsatz I läßt sich in naheliegender Weise auf halbeinfache Gruppen ausdehnen (siehe v. d. Waerden, a.a.O.,[2] S. 785).

BEMERKUNG 3: Für die reellen projektiven Gruppen ist Hauptsatz I bereits von O. Schreier und B. L. van der Waerden bewiesen worden.[4a]

Man wird vermuten, daß die Aussage von Hauptsatz I für keine einfache Gruppe erster Art richtig sei; auf die nicht so ganz leichte Aufzählung aller Isomorphismen für Gruppen erster Art werde ich bald zurückkommen. Ich beabsichtige, dieselben Fragen auch für nicht-halbeinfache Gruppen zu beantworten; für die Gültigkeit von Hauptsatz I ist nämlich etwas wie die Zugehörigkeit zur zweiten Art wichtiger als die Halbeinfachheit.

Den Gedankengang unseres Beweises erkennt man leicht aus der Gliederung der Arbeit in die Sätze 1 — 9. Nur auf Satz 1 und seine wohl an und für sich interessanten Anwendungen will ich besonders hinweisen.

ALLGEMEINE BEZEICHNUNGEN

1. Gruppen: kleine gotische Buchstaben.
Elemente von Gruppen: kleine lateinische Buchstaben.
Gruppen-Eins: $\imath$.
Infinitesimalgruppen: große gotische Buchstaben.
Elemente von Infinitesimalgruppen: große lateinische Buchstaben.
Mengen von Gruppen: kleine fette gotische Buchstaben.
Mengen von Infinitesimalgruppen: große fette gotische Buchstaben.
Meistens werden Gruppe und zugehörige Infinitesimalgruppe mit demselben Buchstaben bezeichnet (klein und groß).
$\cap$ = Durchschnitt, $\cup$ = Vereinigung.

Vorbereitungen—Liesche Gruppen

2. Liesche Gruppe ist bei uns jede Gruppe, die von einer Infinitesimalgruppe erzeugt wird. Wir finden es zweckmäßig, von einer Lieschen Untergruppe *nicht* (wie es manchmal geschieht) Abgeschlossenheit zu verlangen.

Unter einer *Basis* einer Infinitesimalgruppe $\mathfrak{G}$ verstehen wir ein System, aus

[4a] Die Automorphismen der projektiven Gruppen [Abh. Hamburg **6**(1928), 303–322].

dem sich alle Elemente von $\mathfrak{G}$ eindeutig linear kombinieren lassen; ein *Erzeugendensystem* von $\mathfrak{G}$ sei dagegen ein System, aus dem sich alle Elemente von $\mathfrak{G}$ durch lineare Kombination und Kommutatorbildung herstellen lassen.

3. In Lieschen Gruppenkeimen verwendet man öfters geodätische Koordinatensysteme: Ist $X_1, \cdots, X_r$ eine Basis von $\mathfrak{G}$, so erhält der Punkt

$$\exp\left(\textstyle\sum_\nu \lambda^\nu X_\nu\right)$$

die Koordinaten

$$\lambda^1, \cdots, \lambda^r.$$

Alle geodätische Koordinatensysteme transformieren sich ineinander analytisch mit nicht verschwindender Funktionaldeterminante.

Die Koordinatentransformation, bei der der Punkt

$$\exp(\mu^1 X_1 + \cdots + \mu^{\nu_1} X_{\nu_1}) \cdot \exp(\mu^{\nu_1+1} X_{\nu_1+1} + \cdots + \mu^{\nu_2} X_{\nu_2}) \cdots \\ \cdots \exp(\mu^{\nu_p+1} X_{\nu_p+1} + \cdots + \mu^r X_r)$$

die Koordinaten

$$\mu^1, \cdots, \mu^r$$

erhält, ist eine analytische Abbildung des λ-Raums in den μ-Raum mit der Funktionaldeterminante 1 im Nullpunkt.

4. In der Lieschen Gruppe $\mathfrak{g}$ seien $a_\nu(t)$ $(\nu = 1, \cdots, r)$ r stetig differenzierbare Bögen mit $a_\nu(0) = \iota$; die Tangentialvektoren $X_\nu = a'_\nu(0)$ mögen eine Basis von $\mathfrak{G}$ bilden. Dann überdecken die

$$a_1(t_1) \cdots a_r(t_r)$$

eine Umgebung von ι.

Denn die Abbildung des $(t_1, \cdots, t_r)$-Raumes in den λ-Raum, die durch

$$a_1(t_1) \cdots a_r(t_r) = \exp(\lambda^1 X_1 + \cdots + \lambda^r X_r)$$

definiert ist, besitzt in ι die Funktionaldeterminante 1.

5. Reguläre Schichtung einer r-dim. analytischen Mannigfaltigkeit in der Umgebung eines ihrer Punkte nennen wir das analytische Bild (Funktionaldeterminante $\neq 0$) der Schichtung

$$(*) \qquad \xi_{q+1} = \text{const}, \cdots, \xi_r = \text{const}$$

des $(\xi_1, \cdots, \xi_r)$-Raumes; die Bilder der q-dim. Hyperebenen (*) heißen die Schichten.

Wir zeigen: Ein stetig differenzierbarer Bogen, der in jedem seiner Punkte a die durch a laufende Schicht einer regulären Schichtung berührt, verläuft ganz in *einer* Schicht. Es genügt, das zu beweisen für die Schichtung (*) des

ξ-Raumes. In diesem Falle gilt aber, wenn $\xi_\nu(t)$ ($\nu = 1, \cdots, r$) die r Koordinaten des Kurvenpunktes sind, $\xi'_\nu(t) = 0$ für alle $\nu > q$, also in der Tat $\xi_\nu(t) =$ const für alle $\nu > q$, w.z.b.w.

6. Die (Rechts-oder Links-)Nebengruppenschichtung eines Lieschen Gruppenkeimes $\mathfrak{g}$ in bezug auf einen Lieschen Untergruppenkeim $\mathfrak{h}$ ist in der Umgebung jedes Punktes regulär. Anwendung von 5 und des Borelschen Überdeckungssatzes liefert: Eine stetig differenzierbare Kurve in $\mathfrak{g}$, die in jedem ihrer Punkte a die Nebengruppe $a\mathfrak{h}$ berührt, verläuft ganz in *einer* Nebengruppe.

7. In der Lieschen Gruppe $\mathfrak{g}$ seien $c_\nu(t)$ ($\nu = 1, \cdots, n$) analytische Bögen, $c_\nu(0) = \iota$. Die kleinste Gruppe, die diese Bögen enthält, heiße $\mathfrak{h}(\subset \mathfrak{g})$. $\mathfrak{h}$ ist notwendig Liesch.

Beweis: $a(t) = t^p X + t^{p+1}\cdot$(Potenzreihe in t) sei die MacLaurin-Entwicklung eines beliebigen analytischen Bogens aus $\mathfrak{h}$ mit $a(0) = \iota$ und $X \neq 0$; der Vektor X heiße der "Hauptteil" von $a(t)$. $\mathfrak{H}$ sei die Menge aller solcher Hauptteile mit $a(t) < \mathfrak{h}$ nebst dem Nullvektor. $\mathfrak{H}$ ist eine Infinitesimalgruppe, denn sind X bzw. Y die Hauptteile von $a(t)$ bzw. $b(t)$, so sind $\alpha X + \beta Y$ bzw. $[X, Y]$ die Hauptteile der Bögen $a(\alpha t)\cdot b(\beta t)$ bzw. $a(t)b(t)a^{-1}(t)b^{-1}(t)$.

Sei a ein willkürliches Element von $\mathfrak{h}$; dann ist

$$a = c_{\nu_1}(t_1)^{\pm 1} \cdots c_{\nu_q}(t_q)^{\pm 1},$$

wo $\nu_1, \cdots, \nu_q$ gewisse der Zahlen $1, \cdots, n$ (evtl. mit Wiederholungen) sind. Der analytische Bogen

$$a(t) = c_{\nu_1}(tt_1)^{\pm 1} \cdots c_{\nu_q}(tt_q)^{\pm 1} \qquad (0 \leqq t \leqq 1)$$

verbindet ι mit a in $\mathfrak{h}$. Für festes t ist

$$b(\tau) = a(t)^{-1}a(\tau + t) \qquad (-t \leqq \tau \leqq 1 - t)$$

ein analytischer Bogen in $\mathfrak{h}$; nach der Definition von $\mathfrak{H}$ gehört sein Tangentialvektor für $\tau = 0$ zu $\mathfrak{H}$. Also gehört $a'(t)$ (das ist nämlich der Tangentialvektor von $a(\tau + t)$ für $\tau = 0$) zu $a(t)\mathfrak{H}$. Sei $\mathfrak{h}_0$ die von $\mathfrak{H}$ erzeugte (Liesche) Gruppe; nun berührt $a(t)$ in jedem seiner Punkte die Nebengruppe $a(t)\mathfrak{h}_0$, liegt also nach 6 ganz in *einer* Nebengruppe, und da $a(0) = \iota$ ist, ganz in $\mathfrak{h}_0$. Also ist $a = a(1) \in \mathfrak{h}_0$. Da a beliebig in $\mathfrak{h}$ war, haben wir

$$\mathfrak{h} \subset \mathfrak{h}_0. \tag{1}$$

Sei nun $X_1, \cdots, X_r$ eine Basis von $\mathfrak{H}$; definitionsgemäß existieren in $\mathfrak{h}$ analytische Bögen $a_\nu(t)$ mit den Hauptteilen X_ν:

$$a_\nu(t) = t^{p_\nu}X_\nu + \cdots.$$

In $a_\nu(t)$ führen wir als neuen Parameter $s = t^{p_\nu}$ ein,

$$a_\nu(t) = b_\nu(s);$$

$b_\nu(s)$ ist stetig differenzierbar und sein Tangentialvektor für $s = 0$ ist X_ν. Die Menge der

$$b_1(s_1) \cdots b_r(s_r)$$

ist in $\mathfrak{h}$ enthalten und überdeckt nach 4 eine Umgebung von ι in $\mathfrak{h}_0$. Also enthält $\mathfrak{h}$ eine Umgebung von ι in $\mathfrak{h}_0$, also ganz $\mathfrak{h}_0$:

$$\mathfrak{h} \supset \mathfrak{h}_0. \tag{2}$$

Also ((1) und (2)) ist $\mathfrak{h} = \mathfrak{h}_0$ Liesch, w.z.b.w.

8. Satz 1: *In der Untergruppe $\mathfrak{e}$ der Lieschen Gruppe $\mathfrak{g}$ sei jedes Element analytisch verbindbar mit ι. Dann ist $\mathfrak{e}$ Liesch.*

Beweis: $\mathfrak{f}$ sei Liesche Untergruppe maximaler Dimension von $\mathfrak{e}$; $X_1, \cdots, X_p$ sei eine Basis der Infinitesimalgruppe von $\mathfrak{f}$. Sei a beliebig in $\mathfrak{e}$; nach Voraussetzung existiert ein analytischer Bogen $a(t)$ $(0 \leqq t \leqq 1)$ mit $a(0) = \iota$, $a(1) = a$. Nach 7 erzeugen $a(t)$ und $\exp(tX_\nu)$ $(\nu = 1, \cdots, p)$ zusammen eine Liesche Untergruppe von $\mathfrak{g}$, die wegen der Maximalität (s.o.) gleich $\mathfrak{f}$ ist. Also ist $a \in \mathfrak{f}$, also $\mathfrak{e} \subset \mathfrak{f}$, also $\mathfrak{e} = \mathfrak{f}$ Liesch, w.z.b.w.

9. Anwendungen von Satz 1:

1. Sei $\mathfrak{k}$ Teilmenge der Lieschen Gruppe $\mathfrak{g}$; sei $\mathfrak{h}$ eine Teilmenge von $\mathfrak{g}$, die ι enthält, und in der jedes Element mit ι analytisch verbindbar sei (z.B. eine Liesche Untergruppe von $\mathfrak{g}$). Die von den Kommutatoren $k^{-1}h^{-1}kh$ (mit $k \in \mathfrak{k}$ und $h \in \mathfrak{h}$) erzeugte Gruppe $\mathfrak{e}$ ist Liesch.—Denn ein Kommutator $k^{-1}h^{-1}kh$ ist mit ι durch $k^{-1}h(t)^{-1}kh(t)^{-1}$ analytisch verbindbar, wenn $h(t)$ eine analytische Verbindung von ι mit h ist; ist $e \in \mathfrak{e}$, so ist $e = e_1 \cdots e_q$, wo die e_ν von der Gestalt $k^{-1}h^{-1}kh$ sind; ist e_ν durch $e_\nu(t)$ mit ι verbindbar, so auch e durch $e_1(t) \cdots e_q(t)$. Die Voraussetzungen von Satz I sind also erfüllt.
2. Ein Normalteiler $\mathfrak{e}$ der Lieschen Gruppe $\mathfrak{g}$, der mit seiner Kommutatorgruppe identisch ist, ist Liesch.—Man setze nämlich in 9.1: $\mathfrak{h} = \mathfrak{g}$, $\mathfrak{k} = \mathfrak{e}$.
3. Ein Normalteiler $\mathfrak{e}$ der Lieschen Gruppe $\mathfrak{g}$, der nicht ganz im Zentrum von $\mathfrak{g}$ liegt, enthält eine Liesche Untergruppe positiver Dimension.—Ist $a \in \mathfrak{e}$, aber a nicht im Zentrum von $\mathfrak{g}$, so ist die von den $g^{-1}a^{-1}ga$ $(g \in \mathfrak{g})$ erzeugte Gruppe von positiver Dimension und nach Satz I (auch bereits nach 7) Liesch.
4. Ist die Liesche Gruppe $\mathfrak{g}$ einfach im Kleinen, so liegen alle ihre Normalteiler $(\neq \mathfrak{g})$ in ihrem (notwendig diskreten) Zentrum.[5]—Wäre nämlich $\mathfrak{e}$ ein Normalteiler, der nicht ganz im Zentrum läge, so wäre, wie aus Hauptsatz I leicht folgt, die maximale Liesche Untergruppe $\mathfrak{f}$ von $\mathfrak{e}$ ebenfalls Normalteiler von $\mathfrak{g}$ und nach 9.3 von positiver Dimension, was der Einfachheit im Kleinen widerspricht.
5. Man kann 9.4 auf im Kleinen halbeinfache Gruppen ausdehnen.

[5] Verallgemeinerung eines Satzes von v. d. Waerden, a. a. O.,[2] 784, für kompakte Gruppen.

10. Wir erwähnen für spätere Anwendungen den Satz: Bei einer abgeschlossenen Untergruppe einer Lieschen Gruppe ist die Menge der Komponenten (abzählbar) diskret und die Komponente von ι eine Liesche Gruppe.[6]

Vorbereitungen—halbeinfache Gruppen[7]

11. 1. *Rang* l von $\mathfrak{G}$ nennt man bekanntlich die Zahl der algebraisch unabhängigen Wurzeln der Elemente von $\mathfrak{G}$. Bei einer halbeinfachen Gruppe ist der Rang gleich der Zahl der identisch verschwindenden Wurzeln. *Regulär* heißt ein Element von $\mathfrak{G}$, das nicht mehr als l verschwindende Wurzeln besitzt. *Reguläre* Untergruppe der halbeinfachen Gruppe $\mathfrak{G}$ heißt die Gruppe $\mathfrak{H}$ aller mit einem regulären Element vertauschbaren Elemente; auch die von einem regulären $\mathfrak{H}$ erzeugte endliche Gruppe $\mathfrak{h}$ wird regulär genannt.

In Bezug auf eine reguläre Untergruppe $\mathfrak{H}$ (Elemente H) kann man eine halbeinfache Gruppe bekanntlich auf folgende einfache Gestalt bringen:

$$\begin{aligned} [H, E_\alpha] &= \alpha E_\alpha , \\ [E_\alpha , E_{-\alpha}] &= H_\alpha , \\ [E_\alpha , E_\beta] &= N_{\alpha,\beta} E_{\alpha+\beta} (\text{falls } \alpha + \beta \neq 0). \end{aligned}$$

Die $\alpha, \beta, \cdots$ sind dabei lineare Funktionale von $\mathfrak{H}$, die "Wurzelformen"[8] von $\mathfrak{G}$ in bezug auf $\mathfrak{H}$ (die man auch als kontravariante Vektoren im Raume der H auffassen kann). Mit α ist auch $-\alpha$ Wurzelform; unter den Wurzelformen gibt es l linear unabhängige.

2. Wir setzen im Folgenden $\mathfrak{G}$ und $\mathfrak{H}$ stets als reell voraus, gehen aber, wo es nötig ist, ohneweiteres ins Komplexe über. Dann tritt mit α auch $\bar{\alpha}$ (das konjugiert komplexe Funktional) als Wurzelform auf; wir wählen E_α und $E_{\bar{\alpha}}$ so, daß sie konjugiert komplex sind.

Wegen der Halbeinfachheit ist die quadratische Form

$$\begin{aligned} \varphi(T, T) &= \text{Spur der Abbildung } X \to [T[TX]] \\ &= \sum \rho^2 + \sum N_\rho \tau^\rho \tau^{-\rho} \qquad (T = H + \sum \tau^\rho E_\rho) \end{aligned}$$

(Summen über alle Wurzelformen erstrecken!) bekanntlich nicht entartet. N_ρ = Spur der Abbildung $X \to [E_{-\rho}[E_\rho X]]$, also $N_{\bar{\rho}} = \overline{N_\rho}$. Die Normierung[9]

---

[6] E. Cartan, La théorie des groupes finis et continus et l'analysis situs [Mémorial sc. math. XLII], p. 24.

[7] Siehe hierzu: E. Cartan, Sur la structure des groupes de transformations finis et continus [Thèse, Paris 1894].—H. Weyl, Theorie der Darstellungen kontinuierlicher halbeinfacher Gruppen durch lineare Transformationen [Math. Zeitschr. **23**(1925), 271–309; **24** (1926), 328–395], insb. Kap. III.—Wir richten uns hier nach B. L. v. d. Waerden, Die Klassifikation der einfachen Lieschen Gruppen [Math. Zeitschr. **37**(1933), 446–462].

[8] Im Interesse der Deutlichkeit unterscheiden wir zwischen den Wurzeln—das sind die charakteristischen Wurzeln der Matrices der adjungierten Gruppe, entweder für ein einzelnes Element oder als Funktionen auf der ganzen Gruppe—und den Wurzelformen—das sind die Wurzeln, als lineare Funktionen auf $\mathfrak{H}$ definiert.

[9] Siehe v. d. Waerden, a. a. O.,[7] 447.

$N_\rho = -1$ läßt sich daher so durchführen, daß E_ρ und $E_{\bar\rho}$ konjugiert komplex bleiben.

Diese Normierung hat zur Folge, daß H_ρ linear von ρ abhängt. Ist

$$\lambda = \sum_\rho p_\rho \rho$$

eine lineare Kombination von Wurzelformen, so können wir dann widerspruchslos definieren

$$H_\lambda = \sum_\rho p_\rho H_\rho ;$$

alle Elemente von $\mathfrak{H}$ lassen sich auf diese (für uns recht bequeme) Weise darstellen.

Wir haben nun

$$[H_\lambda E_\rho] = -(\rho, \lambda) E_\rho ;$$

$-(\rho, \lambda)$ ist der Wert des Funktionals ρ für den Vektor H_λ. Hier ist (λ, μ) nichts Anderes als die zu der quadratischen Form $Q(H_\lambda, H_\lambda) = \sum \rho^2$ gehörige symmetrische Bilinearform.

Da mit ρ auch $\bar\rho$ Wurzelform ist, ist die Bilinearform (λ, μ) reell (d.h. reell für reelle Argumente). Die quadratische Form (λ, λ) ist im Allgemeinen nicht definit; trotzdem werden wir (mit der nötigen Vorsicht) Bezeichnungen wie orthogonal (in Bezug auf diese quadratische Form) gebrauchen; ein Vektor $H_\lambda \neq 0$ mit $(\lambda, \lambda) = 0$ heißt isotrop; die H_ρ, für die ρ eine Wurzelform ist, können nie isotrop sein, für Wurzelformen ρ gilt: $(\rho, \rho) < 0$.

3. Mit α und β ist auch $\alpha - 2\dfrac{(\alpha, \beta)}{(\beta, \beta)}\beta$ Wurzelform; die $2\dfrac{(\alpha, \beta)}{(\beta, \beta)}$ sind ganz. Mit α, β und $\alpha + k\beta$ ist auch $\alpha + k'\beta$ Wurzelform für $0 \leqq k' \leqq k$.

12. Eine halbeinfache Gruppe ist bekanntlich dann und nur dann komplex einfach, wenn sich je zwei Wurzelformen α, β durch eine "Kette"

$$\alpha = \alpha_1, \qquad \alpha_2, \cdots, \alpha_p, \qquad \alpha_{p+1} = \beta$$

von Wurzelformen verbinden lassen,

$$(\alpha_\nu, \alpha_{\nu+1}) \neq 0.$$

Wir behaupten:

Ist $\mathfrak{G}$ halbeinfach und komplex einfach, und sind α, β Wurzelformen mit $(\alpha, \beta) \neq 0$, so existiert eine Wurzelform γ mit

$$(\alpha, \gamma) \neq 0, \qquad (\beta, \gamma) \neq 0.$$

Zum Beweise nehmen wir an, obige Kette sei minimal; wir müssen dann zeigen, daß $p = 2$ ist. Wir dürfen weiter annehmen

$$\frac{(\alpha_\nu, \alpha_{\nu+1})}{(\alpha_{\nu+1}, \alpha_{\nu+1})} < 0 \quad \text{für alle } \nu$$

(andernfalls ersetzen wir gewisse Wurzelformen durch ihre negativen). Aus der Minimalität folgt

$$(\alpha_\mu, \alpha_\nu) = 0 \text{ für } |\mu - \nu| > 1.$$

Sei bereits für ein $q < p$ bewiesen, daß $\sum_{\nu=2}^{q} \alpha_\nu$ entweder Wurzelform oder 0 ist; wir beweisen dasselbe für $\sum_{\nu=2}^{q+1} \alpha_\nu$. In der Tat ist

$$\frac{\left(\sum_{\nu=2}^{q} \alpha_\nu, \alpha_{q+1}\right)}{(\alpha_{q+1}, \alpha_{q+1})} = \frac{(\alpha_q, \alpha_{q+1})}{(\alpha_{q+1}, \alpha_{q+1})} < 0;$$

ist $\sum_{\nu=2}^{q} \alpha_\nu$ Wurzelform und $\neq -\alpha_{q+1}$, so ist nach 11.3 $\sum_{\nu=2}^{q} \alpha_\nu + \alpha_{q+1}$ Wurzelform; ist $\sum_{\nu=2}^{q} \alpha_\nu = -\alpha_{q+1}$, so ist $\sum_{\nu=2}^{q} \alpha_\nu + \alpha_{q+1} = 0$; ist schließlich $\sum_{\nu=2}^{q} \alpha_\nu = 0$, so ist $\sum_{\nu=2}^{q} \alpha_\nu + \alpha_{q+1} = \alpha_{q+1}$, also Wurzelform. Damit ist die Induktionsbehauptung bewiesen.

Aus ihr folgt: $\gamma = \sum_{\nu=2}^{p} \alpha_\nu$ ist Wurzelform oder 0.

$$(\alpha, \gamma) = (\alpha_1, \gamma) = (\alpha_1, \alpha_2) \neq 0, \qquad (\beta, \gamma) = (\alpha_{p+1}, \gamma) = (\alpha_{p+1}, \alpha_p) \neq 0.$$

γ ist demnach die gesuchte Wurzelform.

13. Aus dem Vorigen folgt: Ist $\mathfrak{G}$ halbeinfach und komplex einfach, so sind alle Zahlen (α, β) (α, β Wurzelformen) rationale Vielfache einer unter ihnen.

14. $\mathfrak{H}$ sei eine p-dim. abelsche Untergruppe der halbeinfachen Gruppe $\mathfrak{G}$; ferner sei $\mathfrak{H}$ von keiner größeren Untergruppe von $\mathfrak{G}$ Normalteiler. Dann ist $\mathfrak{H}$ regulär.

BEWEIS: Nach Voraussetzung gilt für jedes $X \in \mathfrak{G}$ mit

$$[HX] \in \mathfrak{H} \text{ für alle } H \in \mathfrak{H}$$

notwendig

$$X \in \mathfrak{H}.$$

Wir können daher $\mathfrak{G}$ auf die Gestalt

$$\begin{aligned}
&[HE_\alpha] = \alpha E_\alpha \\
&[HE'_\alpha] = \alpha E'_\alpha \bmod E_\alpha \\
&\quad\vdots \\
&[E_\alpha E_{-\alpha}] \in \mathfrak{H}
\end{aligned}$$

bringen, wo die α die nichtverschwindenden Wurzelformen sind. Aus der Tatsache, daß die Spur von

$$X \to [T[TX]]$$

eine nichtentartete quadratische Form ist, schließt man wie üblich, daß $\sum \rho^2$ nicht ausgeartet ist, und daß es genau p linear unabhängige Wurzelformen gibt, also $p \leqq l$ (denn die Wurzeln sind Funktionen auf ganz $\mathfrak{G}$, die Wurzelformen nur auf $\mathfrak{H}$, und für die (linearen) Wurzelformen fallen lineare und algebraische Abhängigkeit zusammen; siehe auch 11.1). Andererseits ist p definitionsgemäß auch die Zahl der verschwindenden Wurzelformen, also $p \geqq l$. Demnach $p = l$; jedes genügend allgemeine Element von $\mathfrak{H}$ ist also regulär, und da $\mathfrak{H}$ sicher maximal abelsch ist, ist auch $\mathfrak{H}$ regulär.

15. Eine Teilmenge der Wurzelformen der halbeinfachen Gruppe nennen wir geschlossen, wenn sie mit α auch $-\alpha$ enthält und mit α und β auch $\alpha + \beta$, falls das eine Wurzelform ist.

Sei **P** eine geschlossene Menge von Wurzelformen und

$$\sigma = \sum_\rho p_\rho \rho$$

eine lineare Kombination aus **P**, die auf ganz **P** orthogonal sei. Dann ist $\sigma = 0$.

Beweis: Die obige Darstellung von σ sei so gewählt, daß $\sum | p_\rho |$ minimal positiv ist. Es gebe ein p_α mit $p_\alpha \neq 0$.

$$(\alpha, \sigma) = \sum p_\rho(\alpha, \rho).$$

Wegen $(\alpha, \sigma) = 0$ und $p_\alpha(\alpha, \alpha) \neq 0$ gibt es ein $\beta(\neq -\alpha)$ mit

$$\operatorname{sgn} p_\alpha(\alpha, \alpha) = -\operatorname{sgn} p_\beta(\alpha, \beta),$$

$$-\operatorname{sgn} \frac{(\alpha, \beta)}{(\alpha, \alpha)} = \operatorname{sgn} \frac{p_\alpha}{p_\beta}.$$

Sei $\operatorname{sgn} p_\alpha = \operatorname{sgn} p_\beta$. Dann ist $\operatorname{sgn} \dfrac{(\alpha, \beta)}{(\alpha, \alpha)} < 0$, also nach 11.3 $\beta + \alpha$ Wurzelform ϵ **P**.

$$\sigma = \sum p_\rho \rho \pm \alpha \pm \beta \mp (\alpha + \beta) = \sum q_\rho \rho;$$

$$q_\alpha = p_\alpha \pm 1,\ q_\beta = p_\beta \pm 1,\ q_{\alpha+\beta} = p_{\alpha+\beta} \mp 1$$

(untere Zeichen für $p_\alpha > 0$, obere für $p_\alpha < 0$). Das ist eine neue Darstellung für σ, und es ist $\sum | q_\rho | < \sum | p_\rho |$. Das ist ein Widerspruch zur Minimalität.

Analog schließt man bei $\operatorname{sgn} p_\alpha = -\operatorname{sgn} p_\beta$.

16. Sei **P** eine geschlossene Menge von Wurzelformen und S das System aller linearen Relationen in **P**. Dann bilden die Relationen der Form

$$(*) \qquad \alpha + \beta = 0 \quad \text{und} \quad \alpha + \beta + \gamma = 0$$

aus S ein Erzeugendensystem für S. ($\alpha, \beta, \gamma \,\epsilon\, \mathbf{P}$.)

Beweis: Das von (*) erzeugte Relationensystem heiße S^*. Ist $S^* \neq S$, so wählen wir die Relation

$$\sum p_\rho \rho = 0$$

in S, aber nicht in S^*, so daß $\sum | p_\rho |$ minimal ist. Wir schließen dann wörtlich wie oben auf die Existenz einer Relation

$$\sum q_\rho \rho = 0$$

in S, aber nicht in S^*, mit

$$\sum | q_\rho | < \sum | p_\rho |,$$

also auf einen Widerspruch.

17. Die Untergruppe $\mathfrak{C}$ der halbeinfachen Gruppe $\mathfrak{G}$ sei von einer Menge von Paaren E_α, $E_{-\alpha}$ aus $\mathfrak{G}$ erzeugt. Dann ist $\mathfrak{C}$ halbeinfach und $\mathfrak{C} \cap \mathfrak{H}$ reguläre Untergruppe von $\mathfrak{C}$.

Beweis: Ein Erzeugendensystem von $\mathfrak{C}$ besteht aus den E_α, $E_{-\alpha}$, H_α, wo α eine geschlossene Menge $\mathbf{A}$ durchläuft. Man berechne $\varphi(T, T)$ für die Gruppe $\mathfrak{C}$. Da N_α ein positives Vielfaches von $(\alpha, \alpha) \neq 0$ ist (siehe Weyl, a.a.O., S.366), brauchen wir uns nur um $\sum \rho^2$ zu kümmern, um zu beweisen, daß $\varphi(T, T)$ nicht ausgeartet ist. Für $\sum \rho^2$ folgt das aber nun aus 15, da danach kein von $\mathbf{A}$ abhängiger Vektor auf ganz $\mathbf{A}$ senkrecht steht.

Rein gruppentheoretische Definition der regulären Untergruppen

Im Großen werden wir die halbeinfachen Gruppen jetzt immer als zentrumfrei voraussetzen.

18. $\mathfrak{h}$ sei reguläre Untergruppe der halbeinfachen Gruppe $\mathfrak{g}$, $\mathfrak{h}^*$ sei die maximal-abelsche Untergruppe von $\mathfrak{g}$ die $\mathfrak{h}$ enthält. Dann ist $\mathfrak{h}^*/\mathfrak{h}$ endlich.[10]

Beweis: Sei $a \in \mathfrak{h}^*$. Der Automorphismus

$$A \cdots x \to a^{-1}xa$$

läßt $\mathfrak{h}$ elementweise fest. A ist auch ein (reeller) Automorphismus für die komplexe Infinitesimalgruppe $\mathfrak{G}$ von $\mathfrak{g}$:

$$[H, E_\alpha] = \alpha E_\alpha, \qquad [H, AE_\alpha] = \alpha AE_\alpha, \qquad \text{also } AE_\alpha = \nu_\alpha E_\alpha.$$

(1) $$[E_\alpha, E_{-\alpha}] = H_\alpha, \qquad [AE_\alpha, AE_{-\alpha}] = H_\alpha, \qquad \text{also } \nu_\alpha \nu_{-\alpha} = 1.$$

(2) $$[E_\alpha, E_\beta] = N_{\alpha,\beta} E_{\alpha+\beta}, \; [AE_\alpha, AE_\beta] = N_{\alpha,\beta} AE_{\alpha+\beta}, \text{ also } \nu_\alpha \nu_\beta = \nu_{\alpha+\beta}, \text{ falls } \alpha + \beta \text{ Wurzelform.}$$

---

[10] In **18** und **19** befinden wir uns auf den Fußspuren von F. Gantmakher, Canonical representations of automorphisms of a complex semi-simple group [Recueil Math. Moscou **5(47)** (1939), 101–146], insb. 128, 129. Da wir mehr mit dem Reellen rechnen müssen, verfahren wir in **18** ein bißchen anders als Gantmakher. Die Resultate von Gantmakher gehen (im Komplexen) weiter als unsere in **18** und **19**; doch spielt das für unsere Untersuchungen keine Rolle.

(3) $\nu_{\bar\alpha}\bar{E}_\alpha = \nu_{\bar\alpha}E_{\bar\alpha} = AE_{\bar\alpha} = A\bar{E}_\alpha = \overline{AE_\alpha} = \overline{\nu_\alpha E_\alpha} = \bar\nu_\alpha\bar{E}_\alpha$, also $\nu_{\bar\alpha} = \overline{\nu_\alpha}$.

(1) und (2) kann man auch so interpretieren: ist $\alpha + \beta = 0$ bzw. $\alpha + \beta + \gamma = 0$ eine Relation zwischen Wurzelformen, so ist $\nu_\alpha\nu_\beta = 1$ bzw. $\nu_\alpha\nu_\beta\nu_\gamma = 1$. Nach 16 folgt hieraus:

$$\text{ist } \sum_\rho p_\rho\rho = 0, \qquad \text{so ist } \prod_\rho \nu_\rho^{p_\rho} = 1 \qquad (p_\rho \text{ ganz}).$$

Auf grund dessen können wir für eine beliebige lineare Kombination

$$\sigma = \sum_\rho q_\rho\rho \qquad (q_\rho \text{ ganz})$$

der Wurzelformen eindeutig definieren

$$\nu_\sigma = \prod_\rho \nu_\rho^{q_\rho}.$$

Diese σ besitzen eine Basis (für ganzzahlige lineare Kombination) $\sigma_1, \cdots, \sigma_l$. Definieren wir

$$\vartheta_{\sigma_\kappa} = \log \nu_{\sigma_\kappa} \text{ (irgendein Wert des log)}$$

und

$$\vartheta_\sigma = \sum_\kappa r_\kappa \log \nu_{\sigma_\kappa} \quad \text{für} \quad \sigma = \sum_\kappa r_\kappa\sigma_\kappa!$$

Dann ist ϑ_σ linear in den σ und für Wurzelformen α

$$e^{\vartheta_\alpha} = \nu_\alpha .$$

Wir definieren weiter

$$\eta_\sigma = \tfrac{1}{2}(\vartheta_\sigma + \overline{\vartheta_{\bar\sigma}}).$$

Dann ist auch η_σ linear in σ, reell für reelle σ und

$$e^{2\eta_\sigma} = \nu_\sigma\overline{\nu_{\bar\sigma}} = \nu_\sigma^2, \qquad e^{\eta_\sigma} = \pm\,\nu_\sigma .$$

Bestimmen wir $H = H_\xi$ so, daß

$$-(\xi\sigma) = \eta_\sigma$$

wird, so ist H reell und erzeugt $h = \exp H$ einen reellen Automorphismus

$$B \cdots x \rightarrow h^{-1}xh$$

mit

$$h^{-1}E_\alpha h = \exp(\eta_\alpha)\cdot E_\alpha = \pm\,\nu_\alpha E_\alpha .$$

Der Automorphismus AB^{-1} von $\mathfrak{g}$ läßt $\mathfrak{h}$ elementweise fest und macht

$$AB^{-1}E_\alpha = \pm E_\alpha .$$

Es gibt nur endlich viel derartige Automorphismen (entsprechend der Zeichenwahl). Da $\mathfrak{g}$ zentrumfrei, a beliebig in $\mathfrak{h}^*$ und $h \in \mathfrak{h}$ ist, besitzt also $\mathfrak{h}$ in $\mathfrak{h}^*$ nur endlich viel Nebengruppen, w.z.b.w.

19. $\mathfrak{h}$ sei reguläre Untergruppe der halbeinfachen Gruppe $\mathfrak{g}$, $\mathfrak{h}'$ sei die größte Untergruppe von $\mathfrak{g}$, in der $\mathfrak{h}$ Normalteiler ist. Dann ist $\mathfrak{h}'/\mathfrak{h}$ endlich.

Beweis: Sei $a \in \mathfrak{h}'$. Der Automorphismus

$$A \cdots x \rightarrow a^{-1}xa$$

läßt $\mathfrak{h}$ als Ganzes fest. Wir setzen

$$AH_\lambda = H_{A\lambda}.$$

A wirkt dann linear auf die λ. Für eine Wurzelform ρ gilt

$$[H_\lambda E_\rho] = -(\rho, \lambda)E_\rho, \quad \text{also} \quad [H_{A\lambda}, AE_\rho] = -(\rho, \lambda)AE_\rho .$$

Also permutiert A die Wurzelformen. Ist A die identische Permutation der Wurzelformen, so ist

$$(\rho, A\lambda) = (\rho, \lambda) \text{ für alle } \rho,$$

also

$$A\lambda = \lambda;$$

dann läßt A also $\mathfrak{h}$ elementweise fest, und dann ist $a \in \mathfrak{h}^*$ (siehe 18). Da die Wurzelformen nur endlich viel Permutationen zulassen, besitzt $\mathfrak{h}^*$ in $\mathfrak{h}'$ nur endlich viel Nebengruppen, also ist $\mathfrak{h}'/\mathfrak{h}^*$ endlich, also nach 18 auch $\mathfrak{h}'/\mathfrak{h}$.

20. 1. Ist $\mathfrak{h}$ eine abelsche Gruppe, so sei $\mathfrak{h}^n$ die Gruppe aller h^n mit $h \in \mathfrak{h}$; der Durchschnitt aller $\mathfrak{h}^n$ ($n = 1, 2, \cdots$) heißt die *Verkürzung* von $\mathfrak{h}$.

2. Man zeigt leicht: Ist $\mathfrak{h}$ abelsch und $\mathfrak{h}_0$, die Komponente von ι in $\mathfrak{h}$, Liesch, ist ferner $\mathfrak{h}/\mathfrak{h}_0$ eine Gruppe von endlich vielen Erzeugenden, so ist $\mathfrak{h}_0$ die Verkürzung von $\mathfrak{h}$.

3. Seien $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ Liesche Untergruppen der Lieschen abelschen Gruppe $\mathfrak{h}$ und $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ die zugehörigen Infinitesimalgruppen. Sei $\mathfrak{C} = \mathfrak{A} \cap \mathfrak{B}$. Dann ist $\mathfrak{c}$ die Verkürzung von $\mathfrak{a} \cap \mathfrak{b}$.

Beweis: Nach 20.2 brauchen wir nur zu zeigen: $(\mathfrak{a} \cap \mathfrak{b})/\mathfrak{c}$ besitzt endlich viel Erzeugende.—Sei

$$h \in \mathfrak{a} \cap \mathfrak{b}.$$

Dann

$$h = \exp H_\lambda, \qquad H_\lambda \in \mathfrak{A} \text{ (reell)},$$

(*)

$$\quad = \exp H_\mu, \qquad H_\mu \in \mathfrak{B} \text{ (reell)};$$

also

(**)

$$(\rho, \lambda) = (\rho, \mu) + n_\rho \cdot 2\pi i \quad (n_\rho \text{ ganz})$$

für alle Wurzelformen ρ. Die Zahlen $n_\rho = n_\rho(h)$ fassen wir zu einem "Vektor" $n(h)$ zusammen. Der Vektor $n(h)$ ist durch h nicht eindeutig bestimmt, sondern nur mod $\mathbf{m}$, der additiven Gruppe aller Vektoren $n(\iota)$. Die $n(h)$ mod $\mathbf{m}$ bilden, wie man leicht sieht, eine additive Darstellung der Gruppe $\mathfrak{a} \cap \mathfrak{b}$. Ist $n(h) \equiv 0$ mod $\mathbf{m}$, so kann man λ und μ in (*) so bestimmen, daß $n(h) = 0$ wird; dann wird in (**) $(\rho, \lambda) = (\rho, \mu)$ für alle ρ, also $\lambda = \mu$, $H_\lambda = H_\mu \in \mathfrak{A} \cap \mathfrak{B} = \mathfrak{C}$ und $h \in \mathfrak{c}$.

Also ist $(\mathfrak{a} \cap \mathfrak{b})/\mathfrak{c}$ homomorphes Bild der (additiven) Gruppe der $n(h)$ mod $\mathbf{m}$, also sicher eine Gruppe von endlich vielen Erzeugenden, w.z.b.w.

21. Folgender Satz definiert rein gruppentheoretisch die regulären Untergruppen:

SATZ 2: *$\mathfrak{h}$ ist dann und nur dann reguläre Untergruppe der halbeinfachen Gruppe $\mathfrak{g}$, wenn gilt:*

1. *$\mathfrak{h}$ ist abelsch;*
2. *$\mathfrak{h}$ ist seine eigene Verkürzung;*
3. *$\mathfrak{h}$ ist maximal in bezug auf* 1 *und* 2 *zusammen;*
4. *ist $\mathfrak{h}_0$ irgendein Normalteiler von $\mathfrak{h}$ und $\mathfrak{h}/\mathfrak{h}_0$ abzählbar, ist weiter $\mathfrak{h}'(\subset \mathfrak{g})$ die größte Gruppe, in der $\mathfrak{h}_0$ Normalteiler ist, so ist auch $\mathfrak{h}'/\mathfrak{h}_0$ abzählbar.*

BEWEIS: Sei $\mathfrak{h}$ regulär. Dann ist 1 trivial, 2 ist in 20.2 und 3 ist in 18 und 20.2 enthalten. Nun zu 4: Sei $\mathfrak{h}/\mathfrak{h}_0$ abzählbar; dann[11] ist $\bar{\mathfrak{h}}_0 = \mathfrak{h}$, also auch $\mathfrak{h}$ Normalteiler von $\mathfrak{h}'$ und nach 19 $\mathfrak{h}'/\mathfrak{h}$ endlich, also $\mathfrak{h}'/\mathfrak{h}_0$ abzählbar, w.z.b.w.

UMGEKEHRT: Mögen Voraussetzungen 1–4 gelten. Sei $a \in \bar{\mathfrak{h}}$. Nach 10 häufen sich die Komponenten von $\bar{\mathfrak{h}}$ nicht; da aber a Häufungspunkt von $\mathfrak{h}$ ist, gibt es ein $b \in \mathfrak{h}$, das in derselben Komponente von $\bar{\mathfrak{h}}$ wie a liegt. Nennen wir die Komponente von $\imath$ in $\bar{\mathfrak{h}}$ nun $\mathfrak{h}_0$, so ist $b^{-1}a = c \in \mathfrak{h}_0$. Es ist $a = bc$, also $\bar{\mathfrak{h}} \subset \mathfrak{h}\cdot\mathfrak{h}_0$. Andererseits $\mathfrak{h} \subset \bar{\mathfrak{h}}$, $\mathfrak{h}_0 \subset \bar{\mathfrak{h}}$, also $\mathfrak{h}\cdot\mathfrak{h}_0 \subset \bar{\mathfrak{h}}$. Demnach $\bar{\mathfrak{h}} = \mathfrak{h}\cdot\mathfrak{h}_0$. Da $\mathfrak{h}$ nach Voraussetzung 1 abelsch ist, da ferner $\mathfrak{h}$ nach Voraussetzung 2 und $\mathfrak{h}_0$ als Liesche abelsche Gruppe (siehe 20.2) seine eigene Verkürzung ist, ist auch $\bar{\mathfrak{h}} = \mathfrak{h}\cdot\mathfrak{h}_0$ seine eigene Verkürzung, also nach Voraussetzung 3: $\mathfrak{h} = \bar{\mathfrak{h}}$. Also ist $\mathfrak{h}$ abgeschlossen.

Sei wieder $\mathfrak{h}_0$ die Komponente von $\imath$ in $\mathfrak{h}$. Dann ist $\mathfrak{h}_0$ Liesch und $\mathfrak{h}/\mathfrak{h}_0$ abzählbar. Wenden wir auf dies $\mathfrak{h}_0$ nun Voraussetzung 4 an! Ist $\mathfrak{h}_0$ Normalteiler in $\mathfrak{h}'$, so muß $\mathfrak{h}'/\mathfrak{h}_0$ abzählbar sein. Insbesondere kann $\mathfrak{h}_0$ nun nicht Normalteiler einer Lieschen Gruppe höherer Dimension sein, ist also nach 14 regulär. Nach 18 (mit $\mathfrak{h}_0$ statt $\mathfrak{h}$ und $\mathfrak{h}$ statt $\mathfrak{h}^*$) und 20.2 ist $\mathfrak{h}_0$ die Verkürzung von $\mathfrak{h}$, also nach Voraussetzung 2: $\mathfrak{h}_0 = \mathfrak{h}$. Also ist $\mathfrak{h}$ regulär, w.z.b.w.

Definition weiterer Untergruppen

Eine feste reguläre Untergruppe $\mathfrak{h}$ werde zugrunde gelegt.

22. 1. $\mathfrak{e}_\alpha$ bzw. $\mathfrak{E}_\alpha$ heiße die von E_α und $E_{\bar{\alpha}}$ erzeugte (reelle) Untergruppe von $\mathfrak{g}$ bzw. $\mathfrak{G}$. Ist $\mathbf{A}$ eine Menge von Wurzelformen, so heiße $\mathfrak{e}_{\mathbf{A}}$ bzw. $\mathfrak{E}_{\mathbf{A}}$ die von der Gesamtheit der $\mathfrak{e}_\alpha$ bzw. $\mathfrak{E}_\alpha$ mit $\alpha \in \mathbf{A}$ erzeugte Untergruppe. Die Menge aller $\mathfrak{E}_\alpha$ heiße $\boldsymbol{\mathfrak{E}}$, die Menge aller $\mathfrak{e}_\alpha$ heiße $\boldsymbol{\mathfrak{e}}$ (α variabel).

2. Sei $\mathfrak{e} \in \boldsymbol{\mathfrak{e}}$ und $\mathfrak{E}$ die zugehörige Infinitesimalgruppe, sei $\mathbf{A}$ die Menge aller α mit $E_\alpha \in \mathfrak{E}$ und $\mathbf{A}^*$ die Menge aller von $\mathbf{A}$ abhängigen Wurzelformen. Wir definieren $\mathfrak{e}^* = \mathfrak{e}_{\mathbf{A}^*}$ ($\mathfrak{e}^*$ hängt von $\mathfrak{e}$ ab); die Infinitesimalgruppe von $\mathfrak{e}^*$ heiße $\mathfrak{E}^*$. Nach 17 ist $\mathfrak{E}^*$ halbeinfach und $\mathfrak{E}^* \cap \mathfrak{H}$ reguläre Untergruppe von $\mathfrak{E}^*$.

[11] Die Überstreichung bedeutet im Augenblick den Übergang zur abgeschlossenen Hülle.

3. Sei $\mathbf{A}$ wie in 2 definiert. Die von den H_α mit $\alpha \in \mathbf{A}$ erzeugte (reelle) Gruppe heiße $\mathfrak{h}_\mathfrak{e}$, die zugehörige Infinitesimalgruppe $\mathfrak{H}_\mathfrak{e}$. Man bemerkt ohneweiteres: $\mathfrak{h}_\mathfrak{e} = \mathfrak{h}_{\mathfrak{e}^*}$.

4. Sei $\mathbf{A}$ wie in 2 definiert. Die von den auf $\mathbf{A}$ *senkrechten* H_λ erzeugte Gruppe heiße $\mathfrak{h}^\mathfrak{e}$, die zugehörige Infinitesimalgruppe $\mathfrak{H}^\mathfrak{e}$. Man sieht ohneweiteres: $\mathfrak{h}^\mathfrak{e} = \mathfrak{h}^{\mathfrak{e}^*}$. Ferner: $\mathfrak{H}^\mathfrak{e}$ und $\mathfrak{H}_\mathfrak{e}$ sind orthogonale Komplemente voneinander.

23. Folgender Satz definiert rein gruppentheoretisch die in 22 eingeführten Untergruppen.

Satz 3: 1. *Die Untergruppe $\mathfrak{e}$ der halbeinfachen Gruppe $\mathfrak{g}$ gehört dann und nur dann zu $\boldsymbol{\mathfrak{e}}$, wenn $\mathfrak{e}$ erzeugt wird von den Kommutatoren $e^{-1}h^{-1}eh$ mit $e \in \mathfrak{e}$ und $h \in \mathfrak{h}$.*

2. $\mathfrak{e}^*$ *ist die größte Gruppe aus $\boldsymbol{\mathfrak{e}}$ mit $\mathfrak{h}^\mathfrak{e} \perp \mathfrak{h}^{\mathfrak{e}^*}$.*

3. $\mathfrak{h}_\mathfrak{e}$ *entsteht durch Verkürzung aus $\mathfrak{e}^* \cap \mathfrak{h}$.*

4. $\mathfrak{h}^\mathfrak{e}$ *entsteht durch Verkürzung aus der Gruppe $(\mathfrak{h}^\mathfrak{e})'$ aller $h \in \mathfrak{h}$ mit $he = eh$ für alle $e \in \mathfrak{e}$.*

Durch Satz 2 ist $\mathfrak{h}$ abstrakt definiert, Satz 3.1 definiert die Gesamtheit der $\mathfrak{e} \in \boldsymbol{\mathfrak{e}}$ abstrakt, Satz 3.4 definiert dann abstrakt zu jedem $\mathfrak{e}$ das $\mathfrak{h}^\mathfrak{e}$, Satz 3.2 definiert abstrakt $\mathfrak{e}^*$ unter Verwendung von $\mathfrak{h}^\mathfrak{e}$, und schließlich wird durch Satz 3.3 $\mathfrak{h}_\mathfrak{e}$ aus $\mathfrak{e}^*$ und $\mathfrak{h}$ definiert.

24. Wir erbringen den Beweis von Satz 3 in der Reihenfolge 1, 4, 2, 3.

1. "Dann": Nach Anwendung 1 von Satz I (siehe 9) ist $\mathfrak{e}$ Liesch. Sei $\mathfrak{E}$ die zugehörige Infinitesimalgruppe; dann ist $[HX] \in \mathfrak{E}$ für alle $H \in \mathfrak{H}$ und $X \in \mathfrak{E}$. Schreiben wir nach 11.1

$$X = H + \sum \tau^\alpha E_\alpha ,$$

so erhalten wir

$$[H_\lambda X] = -\sum \tau^\alpha(\alpha\lambda)E_\alpha \in \mathfrak{E}.$$

Hieraus und aus der Realität von $\mathfrak{E}$ erschließt man leicht, daß $\mathfrak{E}$ von einer Menge von E_α, $E_{\bar\alpha}$ erzeugt wird.

"Nur dann": Wird $\mathfrak{E}$ von einer Menge von E_α, $E_{\bar\alpha}$ erzeugt, so ist es reell und so gilt $[HE] \in \mathfrak{E}$ für alle $E \in \mathfrak{E}$. Daraus folgt die Behauptung.

4. Da $\mathfrak{h}$ abelsch ist, existiert zu jedem $h \in \mathfrak{h}$ ein $H_\lambda \in \mathfrak{H}$ mit $h = \exp H_\lambda$. Es gilt

$$h^{-1}E_\alpha h = \exp(-(\lambda\alpha))\cdot E_\alpha .$$

$h \in (\mathfrak{h}^\mathfrak{e})'$ ist also äquivalent mit

$$\exp(-(\lambda\alpha)) = 1$$

oder

$$(*) \qquad (\lambda\alpha) \equiv 0 \bmod 2\pi i$$

für alle $\alpha \in \mathbf{A}$ (Def. von $\mathbf{A}$, siehe 22.2).

Sei $\mathfrak{K}$ die Gruppe aller (reellen) H_λ mit

$$(\lambda\alpha) = 0 \text{ für alle } \alpha \in \mathbf{A}.$$

Dann ist nach Definition (22.4)

$$(^{**}) \qquad \mathfrak{K} = \mathfrak{H}^{\mathfrak{e}}.$$

Nach (*) erzeugt $\mathfrak{K}$ die Gruppe

$$\mathfrak{k} = \text{Komponente von } \iota \text{ in } (\mathfrak{h}^{\mathfrak{e}})'$$

und ist $(\mathfrak{h}^{\mathfrak{e}})'/\mathfrak{k}$ zyklisch. Nach 20.2 ist die Verkürzung von $(\mathfrak{h}^{\mathfrak{e}})'$ gleich $\mathfrak{k}$, ferner ist nach (**) $\mathfrak{k} = \mathfrak{h}^{\mathfrak{e}}$, also $\mathfrak{h}^{\mathfrak{e}}$ die Verkürzung von $(\mathfrak{h}^{\mathfrak{e}})'$, w.z.b.w.

2. Die Behauptung ergibt sich auf grund des Vorangehenden unmittelbar aus der Tatsache, daß

$$(\alpha\lambda) = 0 \text{ für alle } \alpha \in \mathbf{A}$$

und

$$(\alpha\lambda) = 0 \text{ für alle } \alpha \in \mathbf{A}^*$$

äquivalente Behauptungen sind.

3. Wir setzen $\mathfrak{k} = \mathfrak{e}^* \cap \mathfrak{h}$; $\mathfrak{k}_0$ sei die Komponente von ι in $\mathfrak{k}$, $\mathfrak{K}$ die zugehörige Infinitesimalgruppe. $\mathfrak{K}$ wird erzeugt von den $H_\alpha = [E_\alpha, E_{-\alpha}]$ mit $E_\alpha \in \mathfrak{E}^*$, also $\mathfrak{K} = \mathfrak{H}_{\mathfrak{e}^*} = \mathfrak{H}_{\mathfrak{e}}$, also

$$\mathfrak{k}_0 = \mathfrak{h}_{\mathfrak{e}}.$$

Nach 22.2 (Ende) ist $\mathfrak{k}_0$ reguläre Untergruppe der halbeinfachen Gruppe $\mathfrak{e}^*$, $\mathfrak{k} \supset \mathfrak{k}_0$, $\mathfrak{k}$ ist abelsch, also ist $\mathfrak{k}/\mathfrak{k}_0$ nach 18 endlich, also nach 20.2 $\mathfrak{k}_0$ $(=\mathfrak{h}_{\mathfrak{e}})$ die Verkürzung von $\mathfrak{k}$, w.z.b.w.

Die ausgezeichneten Untergruppen von $\mathfrak{h}$

25. $\mathfrak{M}$ sei die kleinste Menge von Untergruppen von $\mathfrak{H}$ mit der folgenden Eigenschaft:

1. $\mathfrak{M}$ enthält jedes $\mathfrak{H}_{\mathfrak{e}}$ und jedes $\mathfrak{H}^{\mathfrak{e}}$ mit $\mathfrak{e} \in \mathfrak{E}$.
2. $\mathfrak{M}$ enthält mit je zwei Gruppen ihren Durchschnitt und die von ihnen erzeugte Gruppe.

Von selber gilt dann nach 22.4 (und weil Durchschnitt und Erzeugung orthogonal-komplementäre Operationen sind):

3. $\mathfrak{M}$ enthält mit irgendeiner Gruppe auch ihr orthogonales Komplement.

$\mathfrak{e}_\alpha$ war die von E_α und $E_{\bar\alpha}$ erzeugte Gruppe. Statt $\mathfrak{H}^{\mathfrak{e}_\alpha}$ bzw. $\mathfrak{H}_{\mathfrak{e}_\alpha}$ wollen wir kurz $\mathfrak{H}^\alpha$ bzw. $\mathfrak{H}_\alpha$ schreiben. Analog $\mathfrak{h}_\alpha$, $\mathfrak{h}^\alpha$. $\mathfrak{H}_\alpha$ wird erzeugt von H_α und $H_{\bar\alpha}$. Allgemeiner sei (bei beliebigem λ) $\mathfrak{H}_\lambda$ die von H_λ und $H_{\bar\lambda}$ erzeugte Gruppe; ferner $\mathfrak{H}^\lambda$ das orthogonale Komplement von $\mathfrak{H}_\lambda$. Analog $\mathfrak{h}_\lambda$, $\mathfrak{h}^\lambda$.

Wir werden in Zukunft häufig die Fälle unterscheiden:

λ reell (d.h. *rein* reell),

λ imaginär (d.h. *rein* imaginär),

λ komplex (d.h. weder rein reell, noch rein imaginär).
In den beiden ersten Fällen ist $\mathfrak{H}_\lambda$ eindimensional, im dritten Fall zweidimensional (entsprechend $\mathfrak{H}^\lambda$ $(n-1)$- oder $(n-2)$-dimensional).

Wir bemerken schließlich:

4. Enthält $\mathfrak{M}$ eine gewisse nichtisotrope (11.2) eindimensionale Untergruppe $\mathfrak{H}_\lambda$ von $\mathfrak{H}_\alpha$, so enthält es auch noch eine zweite eindimensionale Untergruppe $\mathfrak{H}_\mu$ von $\mathfrak{H}_\alpha$ (nämlich das orthogonale Komplement von $\mathfrak{H}_\lambda$ schneidet $\mathfrak{H}_\alpha$ in einer solchen: $\mu = p\alpha + \bar{p}\bar{\alpha}$, $(\lambda, \mu) = 0$, also wegen der Nichtisotropie $\lambda \neq \mu$).

26. Die Menge $\mathfrak{m}$ (von Untergruppen von $\mathfrak{h}$) sei folgendermaßen definiert: $\mathfrak{k} \in \mathfrak{m}$ dann und nur dann, wenn $\mathfrak{k}$ die von einer Gruppe $\mathfrak{K} \in \mathfrak{M}$ erzeugte Liesche Gruppe ist.

$\mathfrak{m}$ ist dann auch die kleinste Menge von Untergruppen von $\mathfrak{h}$ mit der folgenden Eingenschaft:

1. $\mathfrak{m}$ enthält jedes $\mathfrak{h}_e$ und jedes $\mathfrak{h}^e$ mit $e \in \mathfrak{e}$.
2. $\mathfrak{m}$ enthält mit je zwei Gruppen die Verkürzung ihres Durchschnitts und die von ihnen erzeugte Gruppe. Um das einzusehen, beachte man 20.3.

Nach Satz 3 konnten wir die $\mathfrak{h}_e$ und $\mathfrak{h}^e$ rein gruppentheoretisch definieren. Somit ist auch die Menge $\mathfrak{m}$ (als Funktion von $\mathfrak{h}$ natürlich) rein gruppentheoretisch definiert.

27. Satz 4: *Sei $\mathfrak{g}$ halbeinfach und von zweiter Art. Dann gilt*: 1. *Ist die Gruppe $\mathfrak{h}_e$ eine minimale*[12] *Gruppe aus der Menge $\mathfrak{m}$, so ist $\mathfrak{h}_e$ gleich einem $\mathfrak{h}_\alpha$ und α reelle oder imaginäre Wurzelform.* 2. *Sind alle Wurzelformen komplex, so gibt es Wurzelformen α und β, die auch gleich sein können, derart daß $\mathfrak{h}_{\alpha+\bar{\alpha}}$ und $\mathfrak{h}_{\beta-\bar{\beta}}$ Elemente von $\mathfrak{m}$ sind.*

Beweis: Daß ein minimales $\mathfrak{h}_e$ ein $\mathfrak{h}_\alpha$ ist, ist klar. Die Annahme, α sei komplex, führen wir zum Widerspruch.

Klar ist, daß

$$(\alpha, \bar{\alpha}) = 0$$

sein müßte, da sonst nach 11.3 entweder $\alpha + \bar{\alpha}$ oder $\alpha - \bar{\alpha}$ Wurzelform, also $\mathfrak{h}_{\alpha+\bar{\alpha}}$ oder $\mathfrak{h}_{\alpha-\bar{\alpha}}$ in $\mathfrak{m}$, also $\mathfrak{h}_\alpha$ nicht minimal wäre.

Voraussetzungsgemäß ist $\mathfrak{g}$ komplex-einfach, also gibt es nach 12 eine Wurzelform β mit

$$(\alpha, \beta) \neq 0, \qquad (\bar{\alpha}, \beta) \neq 0.$$

Dann ist auch

$$(\bar{\alpha}, \bar{\beta}) \neq 0, \qquad (\alpha, \bar{\beta}) \neq 0.$$

Wir unterscheiden drei Fälle:

A: β ist reell oder imaginär.—Wegen $(\alpha, \beta) \neq 0$ ist $\mathfrak{H}_\alpha$ nicht senkrecht auf $\mathfrak{H}_\beta$, also (siehe 22.4, Ende) $\mathfrak{H}_\alpha \not\subset \mathfrak{H}^\beta$. Da ferner $\mathfrak{H}^\beta$ $(l-1)$-dim. ist, ist $\mathfrak{H}_\alpha \cap \mathfrak{H}^\beta$

[12] Natürlich minimal bei Vernachlässigung der Gruppe (I).

eindimensional, also eine zu $\mathfrak{M}$ gehörige echte Untergruppe von $\mathfrak{H}_\alpha$. Widerspruch zur Minimalität von $\mathfrak{h}_\alpha$!

B: β ist komplex und

$$(\beta, \bar\beta) \neq 0.$$

Nach 11.3 ist etwa

$$\gamma = \beta + \bar\beta$$

Wurzelform (denn Fall, daß $\beta - \bar\beta$ Wurzelform ist, behandelt man ganz analog), also $\mathfrak{H}_\gamma \in \mathfrak{M}$. Als Wurzelvektor ist H_γ nichtisotrop, also existiert nach 24.4 noch ein $\mathfrak{H}_\delta \in \mathfrak{M}$, $\mathfrak{H}_\delta \subset \mathfrak{H}_\beta$, $\mathfrak{H}_\delta \neq \mathfrak{H}_\gamma$ (δ braucht nicht Wurzelform zu sein). $\mathfrak{H}_\gamma$ und $\mathfrak{H}_\delta$ erzeugen zusammen $\mathfrak{H}_\beta$, und da $(\alpha, \beta) \neq 0$ war, können nicht beide auf $\mathfrak{H}_\alpha$ senkrecht sein. Es gibt demnach ein eindimensionales $\mathfrak{H}_\epsilon$ ($\epsilon = \gamma$ oder δ) in $\mathfrak{M}$, das nicht auf $\mathfrak{H}_\alpha$ senkrecht ist. Das orthogonale Komplement $\mathfrak{H}^\epsilon$ von $\mathfrak{H}_\epsilon$ gehört nach 25.3 zu $\mathfrak{M}$ und schneidet $\mathfrak{H}_\alpha$ eindimensional. $\mathfrak{H}^\epsilon \cap \mathfrak{H}_\alpha \in \mathfrak{M}$ im Widerspruch zur Minimalität von $\mathfrak{h}_\alpha$!

C: β ist komplex und

$$(\beta, \bar\beta) = 0.$$

Wegen $(\alpha, \beta) \neq 0$ ist $\alpha + \beta$ oder $\alpha - \beta$ Wurzelform. Wir dürfen annehmen, $\alpha + \beta$ (sonst ersetzen wir β durch $-\beta$).

Wegen $(\alpha, \bar\alpha) = 0$ haben wir

$$(\alpha + \beta, \bar\alpha) = (\beta, \bar\alpha) \neq 0.$$

Also ist $\alpha + \beta \pm \bar\alpha$ Wurzelform ($\pm$ ist im Folgenden als "$+$ *oder* $-$" zu lesen). Wir unterscheiden zwei Unterfälle:

I. Sei

$$(\alpha + \beta \pm \bar\alpha, \bar\beta) = 0$$

Dann (wegen $(\beta, \bar\beta) = 0$):

$$(*) \qquad \begin{aligned} (\alpha \pm \bar\alpha, \bar\beta) &= 0, \\ (\alpha \pm \bar\alpha, \beta) &= 0. \end{aligned}$$

Also (22.4, Ende)

$$\mathfrak{H}^\beta \supset \mathfrak{H}_{\alpha\pm\bar\alpha}.$$

Andererseits wegen $(\alpha, \beta) \neq 0$

$$\mathfrak{H}^\beta \not\supset \mathfrak{H}_\alpha.$$

Zusammen:

$$(\mathrm{CI}, 1) \qquad \mathfrak{H}^\beta \cap \mathfrak{H}_\alpha = \mathfrak{H}_{\alpha\pm\bar\alpha} \in \mathfrak{M}.$$

Widerspruch zur Minimalität von $\mathfrak{h}_\alpha$.

Wir schließen aber noch weiter, falls alle Wurzelformen komplex sind: $(\alpha + \beta, \overline{\alpha + \beta}) = 0$, also $(\alpha, \bar{\beta}) = -(\bar{\alpha}, \beta)$. Einsetzen in (*) liefert:

$$(\alpha, \beta \mp \bar{\beta}) = 0,$$

$$(\bar{\alpha}, \beta \mp \bar{\beta}) = 0.$$

Hieraus schließt man wie vorhin:

$$\text{(CI, 2)} \qquad \mathfrak{H}^{\alpha} \cap \mathfrak{H}_{\beta} = \mathfrak{H}_{\beta \mp \bar{\beta}} \in \mathfrak{M}.$$

II. Sei nun

$$(\alpha + \beta \pm \bar{\alpha}, \bar{\beta}) \neq 0.$$

Dann hat man vier Möglichkeiten:

$$\begin{array}{ll} \text{a)}\ \alpha + \beta + \bar{\alpha} + \bar{\beta} & \text{Wurzelform} \\ \text{b)}\ \alpha + \beta + \bar{\alpha} - \bar{\beta} & \text{"} \quad \text{"} \\ \text{a}')\ \alpha + \beta - \bar{\alpha} - \bar{\beta} & \text{"} \quad \text{"} \\ \text{b}')\ \alpha + \beta - \bar{\alpha} + \bar{\beta} & \text{"} \quad \text{"}. \end{array}$$

Wir behandeln nur a und b, da a′ und b′ genauso verlaufen.

a) $\gamma = \alpha + \bar{\alpha} + \beta + \bar{\beta}$ ist (reelle) Wurzelform. $\mathfrak{H}_{\gamma} \subset \mathfrak{H}_{\alpha+\beta}$, also existiert nach 25.4 noch ein $\mathfrak{H}_{\delta} \in \mathfrak{M}$, $\mathfrak{H}_{\delta} \subset \mathfrak{H}_{\alpha+\beta}$, $\mathfrak{H}_{\delta} \neq \mathfrak{H}_{\gamma}$. $\mathfrak{H}_{\gamma}$ und $\mathfrak{H}_{\delta}$ erzeugen zusammen $\mathfrak{H}_{\alpha+\beta}$, und da $(\alpha + \beta, \bar{\alpha}) = (\beta, \bar{\alpha}) \neq 0$ war, können nicht beide auf $\mathfrak{H}_{\alpha}$ senkrecht sein. Sind $\mathfrak{H}^{\gamma}$ und $\mathfrak{H}^{\delta}$ die orthogonalen Komplemente, so gilt demnach

$$\mathfrak{H}^{\gamma} \not\supset \mathfrak{H}_{\alpha} \quad \text{oder} \quad \mathfrak{H}^{\delta} \not\supset \mathfrak{H}_{\alpha},$$

also ist $\mathfrak{H}^{\gamma} \cap \mathfrak{H}_{\alpha}$ oder $\mathfrak{H}^{\delta} \cap \mathfrak{H}_{\alpha}$ eindimensional und in $\mathfrak{M}$ (siehe 25.4) im Widerspruch zur Minimalität von $\mathfrak{h}_{\alpha}$.

b) $\gamma = \alpha + \bar{\alpha} + \beta - \bar{\beta}$ ist Wurzelform, $2\mathfrak{R}\alpha = \mathfrak{R}\gamma$, $2\mathfrak{J}\beta = \mathfrak{J}\gamma$. $\mathfrak{H}_{\gamma} \cap \mathfrak{H}_{\alpha} = \mathfrak{H}_{\alpha+\bar{\alpha}} \in \mathfrak{M}$, $\mathfrak{H}_{\gamma} \cap \mathfrak{H}_{\beta} = \mathfrak{H}_{\beta-\bar{\beta}} \in \mathfrak{M}$.—Widerspruch zur Minimalität von $\mathfrak{h}_{\alpha}$.

Damit ist die erste Hälfte des Satzes bewiesen. Die Voraussetzung der zweiten Hälfte (alle Wurzelformen komplex) hat auch wieder $(\alpha, \bar{\alpha}) = 0$ zur Folge; sie hat weiter zur Folge, daß wir uns in den Fällen CI oder CIIb befinden —in beiden Fällen haben wir aber die zweite Behauptung bereits bewiesen (siehe die Formeln CI, 1 und CI, 2 und die Formeln in CIIb).

28. Ist

$$(\alpha, \bar{\alpha}) = 0,$$

so ist

$$[H'_{\alpha}, E'_{\alpha}] = \vartheta E'_{\alpha} \qquad (\vartheta \text{ reell}),$$

falls

$$E'_{\alpha} = E_{\alpha} + E_{\bar{\alpha}}, \qquad H'_{\alpha} = H_{\alpha} + H_{\bar{\alpha}}, \qquad \vartheta = (\alpha, \alpha).$$

ist.

Rein gruppentheoretische Definition kompakter Untergruppenkeime

29. Satz 5: *$\mathfrak{g}$ sei halbeinfach von zweiter Art. Man wähle, wenn möglich, ein $\mathfrak{h}' = \mathfrak{h}_\mathfrak{e}$, so daß es minimal in $\mathfrak{m}$ ist, und dabei $\mathfrak{e} \epsilon \mathfrak{E}$ (auch $\mathfrak{e}'$ genannt) so, daß es auch minimal ist. Ist das nicht möglich, so wähle man $\mathfrak{e} \epsilon \mathfrak{E}$ und eine in $\mathfrak{m}$ minimale Untergruppe $\mathfrak{h}'$ von $\mathfrak{h}_\mathfrak{e}$, so daß gilt:*

1. *$\mathfrak{e}$ ist minimal;*
2. *es gibt eine echte Untergruppe $\mathfrak{e}' \neq (\imath)$ von $\mathfrak{e}$, derart daß die Kommutatoren $h^{-1}e^{-1}he$ $(h \epsilon \mathfrak{h}',\ e \epsilon \mathfrak{e}')$ die Gruppe $\mathfrak{e}'$ erzeugen.*

Diese Wahl ist möglich, und es gilt weiter:

A. *Entweder ist $\mathfrak{h}'$ eine kompakte Gruppe.*

B. *Oder $\mathfrak{h}'$ ist nicht kompakt,*

$$\mathfrak{h}' = (\exp tH'), \qquad \mathfrak{e}' = (\exp tE'),$$

$$[H'E'] = \vartheta E' \qquad (\vartheta \text{ reell}).$$

Die beiden Fälle sind noch abstrakt dadurch unterschieden, daß in A *$\mathfrak{h}'$ Elemente endlicher Ordnung besitzt, in* B *nicht.*

30. Beweis von Satz 5: Sind nicht alle Wurzelformen komplex, so ist $\mathfrak{h}'$ wegen der Minimalität und nach Satz 4 ein $\mathfrak{h}_\alpha$ mit reellem oder imaginärem α; $\mathfrak{e}' = \mathfrak{e}_\alpha$ oder $\mathfrak{e}_{-\alpha}$. Ist α reell, so befinden wir uns im Fall B, $\vartheta = (\alpha, \alpha)$. Sei α imaginär. Dann wird $\mathfrak{e}_\alpha$ erzeugt von E_α, $E_{-\alpha}$. Eine Basis von $\mathfrak{E}_\alpha$ ist E_α, $E_{-\alpha}$, H_α, und es gilt

$$[E_\alpha, E_{-\alpha}] = H_\alpha,$$

also nach Übergang zum Konjugierten

$$[E_{-\alpha}, E_\alpha] = \bar{H}_\alpha,$$

also $\bar{H}_\alpha = -H_\alpha$, also H_α imaginär. Wegen $(\alpha, \alpha) < 0$ und 13 ist auch (α, β) reell für alle Wurzelformen α, β. Ferner ist

$$H' = iH_\alpha$$

reell und

$$\mathfrak{h}' = (\exp tH')_{t \text{ reell}}.$$

In

$$[H'E_\beta] = -i(\alpha, \beta)E_\beta$$

sind wegen der Realität der (α, β) die Koeffizienten $-i(\alpha, \beta)$ imaginär. Also sind alle Wurzeln von H' imaginär und rationale Vielfache voneinander, und demnach ist $\mathfrak{h}'$ kompakt (Fall A).

Sind hingegen alle Wurzelformen komplex, so versagt nach Satz 4, 1 die erste Bestimmungsmethode für $\mathfrak{h}'$. Nach Satz 4, 2 und nach 28 ist die vorgeschriebene Wahl von $\mathfrak{h}'$ und $\mathfrak{e}'$ möglich. Sei diese Wahl irgendwie geschehen. Dann ist $\mathfrak{h}'$ eindimensional nach Satz 4, 1 und $\mathfrak{e}'$ Liesch nach Satz 1 und eindimensional als echte Untergruppe von $\mathfrak{e}$. Wir befinden uns somit wieder im Fall B.

31. Satz 6: *Ein eindimensionaler kompakter Untergruppenkeim der halbeinfachen Gruppe zweiter Art* $\mathfrak{g}$ *läßt sich rein gruppentheoretisch beschreiben.*

Beschreibung: Im Falle A sind wir nach Satz 5 bereits fertig.—Fall B:

$$[H'E'] = \vartheta E',$$

$$h_t = \exp\left(\frac{t}{\vartheta} H'\right), \qquad e_s = \exp(sE'),$$

$$h_t E' h_t^{-1} = (\exp t) \cdot E'$$

$$h_t e_s h_t^{-1} = e_{s \cdot \exp t}\,. \tag{*}$$

Sei $e_{s_0}(\neq \imath)$ beliebig in $\mathfrak{e}'$; die Menge aller $h_t e_{s_0} h_t^{-1}$ nennen wir einen Strahl. Nach (*) gibt es in $\mathfrak{e}'$ nur die zwei Strahlen e_τ, $\tau > 0$, und e_τ, $\tau < 0$. Sei $\mathfrak{s}$ ein Strahl von $\mathfrak{e}'$, $e_{s_0} \in \mathfrak{s}$ und $\mathfrak{t}$ der andere Strahl.

$$\mathfrak{u} = \mathfrak{s} \cap e_{s_0} \cdot \mathfrak{t}$$

ist die Menge aller e_s mit s zwischen 0 und s_0.

$$\mathfrak{u} \cup \mathfrak{u}^{-1} \cup (e_{s_0}) \cup (e_{s_0}^{-1}) \cup (\imath)$$

ist also ein kompakter Keim von $\mathfrak{e}'$ und rein gruppentheoretisch beschrieben.

Rein gruppentheoretische Beschreibung einer kompakten Umgebung der Eins

32. Die Elemente

$$X_1, \cdots, X_{p_1} \qquad (p_1 > 1)$$

mögen die r-parametrige Gruppe $\mathfrak{G}$ erzeugen (Halbeinfachheit brauchen wir hier nicht vorauszusetzen). Man erweitere dies System zu einem System

$$X_1, \cdots, X_{p_1}, \cdots, X_{p_2}$$

durch Hinzufügung der Kommutatoren

$$[X_\rho X_\sigma] \qquad (\rho, \sigma \leqq p_1),$$

dies analog zu einem System

$$X_1, \cdots, X_{p_1}, \cdots, X_{p_2}, \cdots, X_{p_3}$$

durch Hinzufügung der Kommutatoren

$$[X_\rho X_\sigma] \qquad (\rho, \sigma \leqq p_2),$$

und fahre sofort. Dann gibt es ein u, das nur von r abhängt, derart daß

$$X_1, \cdots, X_u$$

eine Basis von $\mathfrak{G}$ bilden.

Man nehme nämlich $u = p_{r-1}$. Ist v so bestimmt, daß die

$$X_1, \cdots, X_{p_v}, \cdots, X_{p_{v+1}} \qquad (\mathrm{X}_{v+1})$$

von den

$$X_1, \cdots, X_{p_v} \qquad (\mathbf{X}_v)$$

linear abhängen, so enthält $\mathbf{X}_v$ eine Basis von $\mathfrak{G}$. Sei v minimal. Dann liegt für $\nu < v$ in

$$X_1, \cdots, X_{p_\nu}, \cdots, X_{p_\nu+1}$$

mindestens ein linear unabhängiges Element mehr als in

$$X_1, \cdots, X_{p_\nu}.$$

Da es in $\mathfrak{G}$ nur r unabhängige Elemente gibt, muß also $v < r$ sein. Somit reicht $u = p_{r-1}$ in der Tat aus.

33. Satz 7: *In der einfachen Gruppe zweiter Art* $\mathfrak{g}$ *läßt sich eine kompakte Teilmenge* U*, die* $\imath$ *als inneren Punkt enthält, rein gruppentheoretisch beschreiben.*

Beschreibung: $\mathfrak{k}$ sei der kompakte eindimensionale Untergruppenkeim aus Satz 6. Wegen der Einfachheit von $\mathfrak{g}$ kann man $a_1, \cdots, a_p$ so wählen, daß die ebenfalls kompakten Keime

$$\mathfrak{k}_\nu = a_\nu^{-1}\mathfrak{k}a_\nu \qquad (\nu = 1, \cdots, p)$$

$\mathfrak{g}$ erzeugen;

$$\mathfrak{k}_\nu : k_\nu(s) \qquad (0 \leqq s \leqq 1).$$

Zu den $k_\nu(s)$ füge man die Kommutatorkurven

$$k_\rho(s)^{-1}k_\sigma(s)^{-1}k_\rho(s)k_\sigma(s) \qquad (\rho < \sigma),$$

so daß das System

$$k_1(s), \cdots, k_{p_1}(s), \cdots, k_{p_2}(s) \qquad (p_1 = p)$$

entsteht. So fahre man fort bis zum Index u aus 32. Nach 32 kommt unter den Hauptteilen von

$$k_1(s), \cdots, k_u(s)$$

eine Basis von $\mathfrak{G}$ vor. Nach 4 enthält die offenbar kompakte Menge

$$U : k_1(s_1) \cdots k_u(s_u) \qquad (0 \leqq s_\nu \leqq 1)$$

eine Umgebung von $\imath$, wie wir es wünschten.

Beweis des Hauptsatzes I

34. Satz 8: *In der einfachen Gruppe zweiter Art* $\mathfrak{g}$ *lassen sich beliebig kleine kompakte Teilmengen* V*, die* $\imath$ *als inneren Punkt enthalten, rein gruppentheoretisch beschreiben.*

Beschreibung: a sei beliebig $\neq \imath$. U sei gemäß Satz 7 gewählt.

$$V = \bigcup_{\substack{b_\nu \in U \\ c_\nu \in U}} \prod_{\nu=1}^{r} c_\nu^{-1}(b_\nu^{-1}a^{-1}b_\nu a)c_\nu.$$

Wörtlich wie bei v. d. Waerden, a.a.O.,[2] 783–784, zeigt man, daß die (von a abhängenden) V die gewünschten Eigenschaften besitzen. Nur verwendet man hier statt der Kompaktheit von $\mathfrak{g}$ die von U (Satz 7).

35. Der Satz 8 ist mit unserm Hauptsatz I äquivalent (siehe dazu auch v. d. Waerden, a.a.O.,[2] 784–785), wenn wir *beide* miteinander isomorphe Gruppen als von zweiter Art annehmen. Da die Einfachheit von selber ein rein gruppentheoretischer Begriff ist, haben wir nur noch zu zeigen, daß auch die "zweite Art" oder—was auf dasselbe hinauskommt—die "erste Art" rein gruppentheoretisch zu beschreiben ist, um den Hauptsatz I ganz bewiesen zu haben. Das geschieht nun:

Rein gruppentheoretische Beschreibung der "ersten" Art

36. Satz 4′: *Eine halbeinfache Gruppe ist dann und nur dann von erster Art, wenn für je zwei minimale Gruppen* $\mathfrak{e}$ *und* $\mathfrak{e}'$ *aus* $\mathfrak{E}$ *gilt:*

1) $\mathfrak{e}$ *ist abelsch;*
2) $\mathfrak{e}$ *enthält Elemente endlicher Ordnung;*
3) *Verkürzung von* $\mathfrak{h}_{\mathfrak{e}} \cap \mathfrak{h}^{\mathfrak{e}'} = (\imath)$ *oder* $\mathfrak{h}_{\mathfrak{e}}$;
4) *Verkürzung von* $\mathfrak{h}_{\mathfrak{e}} \cap \mathfrak{h}_{\mathfrak{e}'} = (\imath)$ *oder* $\mathfrak{h}_{\mathfrak{e}}$.

Beweis: "Nur dann": Die Wurzelformen einer Gruppe erster Art zerfallen in zwei Systeme $\mathbf{P}$ und $\bar{\mathbf{P}}$. Ist $\alpha \in \mathbf{P}$, so ist $\bar{\alpha} \in \bar{\mathbf{P}}$ und umgekehrt; ist α, $\beta \in \mathbf{P}$, so ist $(\alpha, \bar{\beta}) = 0$.

Sei $\mathfrak{g}$ von erster Art. Jedes minimale $\mathfrak{e}$ ist von der Gestalt $\mathfrak{e}_\alpha$. Das zugehörige $\mathfrak{E}_\alpha$ wird von E_α und $E_{\bar{\alpha}}$ erzeugt. Wegen $(\alpha, \bar{\alpha}) = 0$ ist $[E_\alpha E_{\bar{\alpha}}] = 0$, also 1 erfüllt.

Seien α, $\rho \in \mathbf{P}$. Man bilde

$$K = i(\bar{\alpha}, \bar{\alpha})H_\alpha - i(\alpha, \alpha)H_{\bar{\alpha}};$$

aus $(\alpha, \bar{\rho}) = (\bar{\alpha}, \rho) = 0$ folgt

$$[KE_\rho] = -i(\bar{\alpha}, \bar{\alpha})(\alpha, \rho)E_\rho,$$

$$[KE_{\bar{\rho}}] = \quad i(\alpha, \alpha)(\bar{\alpha}, \bar{\rho})E_{\bar{\rho}}.$$

Aus 11.3 und $(\alpha, \alpha) = (\bar{\alpha}, \bar{\alpha}) < 0$ schließt man, daß die Koeffizienten rechts (für alle $\rho \in \mathbf{P}$), d.h. die Wurzeln von K, rein imaginär und rationale Vielfache voneinander sind. Die von K erzeugte Untergruppe $\mathfrak{k}$ von $\mathfrak{h}_{\mathfrak{e}}$ ist also kompakt, und daraus folgt 2.

$\mathfrak{e} = \mathfrak{e}_\alpha$, $\mathfrak{e}' = \mathfrak{e}_\beta$. Sei $H_\lambda \in \mathfrak{H}_\alpha \cap \mathfrak{H}^\beta$. Dann

$$\lambda = p\alpha + \bar{p}\bar{\alpha}, \qquad (\lambda, \beta) = 0,$$

also wegen $(\bar{\alpha}, \beta) = 0$

$$p(\alpha, \beta) = 0.$$

Also entweder $(\alpha, \beta) = 0$ und dann $\mathfrak{H}_\alpha$ senkrecht auf $\mathfrak{H}_\beta$ und $\mathfrak{H}_\alpha \cap \mathfrak{H}^\beta = \mathfrak{H}_\alpha$. Oder $p = 0$, $\bar{p} = 0$, $\lambda = 0$ und dann $\mathfrak{H}_\alpha \cap \mathfrak{H}^\beta = (0)$. Also gilt 3 (siehe 20.3).

Wir betrachten nun $H_\lambda \in \mathfrak{H}_\alpha \cap \mathfrak{H}_\beta$.

$$\lambda = p\alpha + \bar{p}\bar{\alpha} = q\beta + \bar{q}\bar{\beta}$$

"Multiplikation" mit ρ liefert:

$$p(\alpha\rho) = q(\beta\rho) \qquad \text{für alle } \rho,$$

also

$$p\alpha = q\beta.$$

Also entweder $p = q = 0$, $\lambda = 0$ und dann $\mathfrak{H}_\alpha \cap \mathfrak{H}_\beta = (0)$. Oder $p \neq 0$ und dann $\mathfrak{H}_\alpha = \mathfrak{H}_\beta$. Also gilt 4 (siehe 20.3).

"Dann": Wegen Voraussetzung 1 kann das α eines minimalen $\mathfrak{e}_\alpha$ nicht imaginär sein ($[E_\alpha E_{-\alpha}]$ ist doch $= H_\alpha$), wegen Voraussetzung 2 kann es nicht reell sein ($\mathfrak{e}_\alpha$ ist dann offen); also ist für ein minimales $\mathfrak{e}_\alpha$ das α komplex. Da andererseits $\mathfrak{e}_\alpha$ für reelles oder imaginäres α sicher minimal ist, müssen alle Wurzelformen komplex sein. Wäre $\mathfrak{g}$ von zweiter Art, so kämen nur die Fälle CI und CIIb aus 27 in Frage. CI ist ausgeschlossen wegen $\mathfrak{H}^\alpha \cap \mathfrak{H}_\beta \neq \mathfrak{H}_\beta$, $\neq (0)$ (Formel CI, 1) und CIIb wegen $\mathfrak{H}_\gamma \cap \mathfrak{H}_\alpha \neq \mathfrak{H}_\alpha$, $\neq (0)$. Also ist $\mathfrak{g}$ von erster Art.

Einfache Gruppen mit diskretem Zentrum

37. Wir beweisen die Bemerkung 1 zu Hauptsatz I.

Die einfachen Gruppen zweiter Art $\mathfrak{f}$ und $\mathfrak{g}$ dürfen ein diskretes Zentrum besitzen. A sei ein Isomorphimus, $A\mathfrak{f} = \mathfrak{g}$. Die resp. Faktorgruppen nach dem Zentrum seien $\mathfrak{f}'$ und $\mathfrak{g}'$; φ und ψ seien die zugehörigen Homomorphismen, $\varphi\mathfrak{f} = \mathfrak{f}'$, $\psi\mathfrak{g} = \mathfrak{g}'$. A induziert einen Isomorphismus A', $A'\mathfrak{f}' = \mathfrak{g}'$,

$$A'\varphi = \psi A. \tag{1}$$

Die universellen Überlagerungsgruppen von $\mathfrak{f}$ und $\mathfrak{g}$ seien $\tilde{\mathfrak{f}}$ und $\tilde{\mathfrak{g}}$, die zugehörigen Projektionen Φ und Ψ, $\Phi\tilde{\mathfrak{f}} = \mathfrak{f}$, $\Psi\tilde{\mathfrak{g}} = \mathfrak{g}$. A' ist nach Hauptsatz I stetig und induziert, wie man weiß, einen stetigen Homomorphismus $\tilde{A}$, $\tilde{A}\tilde{\mathfrak{f}} = \tilde{\mathfrak{g}}$,

$$A'\varphi\Phi = \psi\Psi\tilde{A}. \tag{2}$$

Wir definieren

$$\begin{aligned} A\Phi &= B, \\ \Psi\tilde{A} &= C. \end{aligned} \tag{3}$$

Aus (1), (2), (3) folgt

$$\psi B = \psi A\Phi = A'\varphi\Phi = \psi\Psi\tilde{A} = \psi C. \tag{4}$$

Das ψ-Urbild von $\imath$ ist das Zentrum von $\mathfrak{g}$, also folgt aus (4):

Für jedes $x \in \tilde{\mathfrak{f}}$ ist $(Bx)\cdot(Cx)^{-1}$ im Zentrum von $\mathfrak{g}$. Da das Zentrum abzählbar ist, kann der Normalteiler $\mathfrak{d}$ aller x mit

$$Bx = Cx$$

nicht abzählbar sein, sicher also nicht im Zentrum von $\mathfrak{f}$ liegen. Nach Anwendung 4 von Satz 1 ist dann $\mathfrak{d} = \mathfrak{f}$, also

$$Bx = Cx$$

für alle x. Da $\tilde{A}$, also nach (3) auch C stetig war, ist B stetig. Da Φ als im Kleinen topologische Abbildung beiderseits stetig ist (Bilder offener Mengen sind offen), ist nach (3) auch A stetig, w.z.b.w.

AMSTERDAM, HOLLAND

Annals of Mathematics
Vol. 42, No. 5, December, 1941

A TAUBERIAN THEOREM FOR PARTITIONS

By A. E. Ingham

(Received August 11, 1940)

1. Introduction

The formulae

$$p(n) \sim \frac{e^{\pi\sqrt{(2n/3)}}}{(4\sqrt{3})n}, \qquad q(n) \sim \frac{e^{\pi\sqrt{(n/3)}}}{4\cdot 3^{\frac{1}{4}}\cdot n^{\frac{3}{4}}}, \qquad (n \to \infty), \tag{1}$$

where $p(n)$ is the number of unrestricted partitions of n and $q(n)$ the number of partitions of n into unequal parts (or into odd parts), were discovered by Hardy and Ramanujan [4], and independently by Uspensky [7]. Their proofs give much more than simple asymptotic equality, but demand a rather special study of the generating functions

$$\sum_{n=0}^{\infty} p(n)z^n = \prod_{\nu=1}^{\infty} \frac{1}{1 - z^\nu},$$

$$\sum_{n=0}^{\infty} q(n)z^n = \prod_{\nu=1}^{\infty} (1 + z^\nu) = \prod_{\nu=1}^{\infty} \frac{1}{1 - z^{2\nu-1}},$$

based on the transformation theory of the elliptic modular functions. Later researches have tended in the direction of a still deeper study of particular problems, culminating in the exact formulae of Rademacher [6] and his followers, or in the direction of a broader and more elementary treatment giving less precise results than (1). But a specification of properties of the generating functions sufficient for the deduction of (1), but not highly extravagant, seems to be lacking.

In this paper we show how to deduce (1) from an elementary knowledge of the asymptotic behaviour of the generating functions when $z(= x + iy)$ approaches the principal singularity $z = 1$ in an arbitrarily wide 'Stolz angle' $|y| \leqq \Delta(1 - x)$, $(0 < \Delta < \infty)$. A general result to which the method naturally leads is set out as Theorem 2 in §3, with indications of some of the more immediate applications.

Theorem 2 is deduced from Theorem 1, a Tauberian theorem for the integral

$$f(s) = \int_0^\infty e^{-us}\, dA(u), \qquad A(0) = 0, \qquad (s = \sigma + it), \tag{2}$$

which occupies an intermediate position between theorems of Hardy-Littlewood type with conclusion $A(u) \sim u^a$ and of Wiener-Ikehara type with conclusion $A(u) \sim e^{au}$, $(u \to \infty;\ a > 0)$. The nature of the interpolation is exhibited in

the following table, in which a simple specimen of each type of theorem has been selected.

HYPOTHESES		CONCLUSION
(A_1) $f(s) \sim 1/s$ when $s \to 0$ by real positive values;	$A(u)$ increasing;	$A(u) \sim u$.
(A_2) $f(s) \sim \left(\dfrac{2\pi s}{1-\alpha}\right)^{\frac{1}{2}} \exp\left\{\dfrac{1-\alpha}{\alpha}\left(\dfrac{1}{s}\right)^{\alpha/(1-\alpha)}\right\}$ when $s \to 0$ in every fixed 'Stolz angle' $\lvert t \rvert \leqq \Delta\sigma$, $(0 < \Delta < \infty)$;		$A(u) \sim u^{-\frac{1}{2}} e^{u^{\alpha}/\alpha}$, $(0 < \alpha < 1)$.
(A_3) $f(s) - 1/(s-1) \to$ limit when $\sigma \to 1 + 0$, uniformly in every fixed interval $\lvert t \rvert \leqq T$, $(0 < T < \infty)$;		$A(u) \sim e^u$.

Theorem 1, of which (A_2) is a very special case, is stated and proved, on the lines of the Heilbronn-Landau version of the Wiener-Ikehara theorem, in §2.

It will be seen that, in the method of interpolation shown in the above table, the region involved in the hypothesis on $f(s)$ undergoes a discontinuous change of form at each end of the range but no steady change in the interior. This is in contrast to some other possible methods, of which two may be mentioned here. In one method, essentially that of Wiener and Pitt [8], we first transform (A_1) and (A_3) into equivalent theorems (A_1^*) and (A_3^*) by changing '$f(s) - (s-1)^{-1} \to$ limit when $\sigma \to 1 + 0, \cdots$,' to '$f(s) - s^{-1} \to$ limit when $\sigma \to +0, \cdots$,' in the hypotheses of (A_3) and replacing the conclusions by

$$\frac{A(u + \delta h_u) - A(u - \delta h_u)}{2\delta h_u} \to 1 \quad \text{as} \quad u \to \infty \quad \text{for fixed } \delta, \qquad (0 < \delta \leqq 1),$$

where h_u is u in (A_1^*) and 1 in (A_3^*);[1] and we then interpolate by taking intermediate forms of h_u, say $h_u = u^{\vartheta}$, $(1 > \vartheta > 0)$. Another method, outlined by Avakumović [1], is to interpolate similarly (by way of h_u) between the convergence theorems associated with (A_1) and (A_3), in which the conclusion is $A(u) \to A$ and the hypothesis on $A(u)$, taken in the 'Schmidt' form, is

$$\varliminf_{\substack{u,u' \to \infty \\ u \leqq u' \leqq u + \delta h_u}} \{A(u) - A(u')\} \geqq -\epsilon(\delta), \qquad \lim_{\delta \to +0} \epsilon(\delta) = 0,$$

with h_u as before. In these two methods the hypotheses on $f(s)$ (which need not be specified in detail here) involve a region of the form $\lvert t \rvert \leqq C\sigma^{\vartheta}$ which gradually extends (in the neighbourhood of $s = 0$) as ϑ decreases. This extension of the region is a natural way of tightening the bond between $f(s)$ and its approximation ($1/s$ or A) to balance the strengthening of the conclusion, or the weakening of the hypothesis on $A(u)$, as the order of h_u decreases. But the fact

[1] The equivalence of (A_1) and (A_1^*) is trivial, and that of (A_3) and (A_3^*) is easily proved by setting up the relation $f(s) = f^*(s-1)$, where $f(s)$ and $f^*(s)$ satisfy the conditions of (A_3) and (A_3^*), respectively.

that the boundary of the region touches the imaginary axis at $s = 0$ makes this kind of hypothesis more difficult to verify in applications than the 'Stolz angle' hypotheses used in this paper.

An explicit deduction of the asymptotic formula for $p(n)$ from a Tauberian theorem has been indicated by Avakumović [1, 2], but his demands on the generating function are much more severe, as regards both the region and the degree of the approximation, than those made by Theorem 1.[2]

There is also a paper by Martin and Wiener [5] which works in broadly the same range as Theorem 1 with regard to the rate of growth of $f(s)$ with $1/s$, but on a real variable basis A passage to the complex domain is effected by means of Parseval's theorem in the case of power series, and the possibility of arithmetical applications is announced. This paper, and a paper by Avakumović and Karamata [3], will be referred to again in §4 in the course of some comments on the conditions of Theorem 1.

I am indebted to Professor Wiener for his help in elucidating the relationship between various methods and results.

NOTATION. We write $f \prec g$, $f \succ g$, $f \preccurlyeq g$ for $f/g \to 0$, $f/g \to \infty$, $f = O(g)$ respectively, where f and g are positive functions.

A sloping arrow indicates monotonic approach to a (finite or infinite) limit.

We denote the open and closed segments with end points a and b in the complex plane by (a, b) and $[a, b]$, respectively, and use the mixed notations $(a, b]$ and $[a, b)$ with the obvious meanings.

Limit operations in the complex plane are, of course, to be understood in the usual two-dimensional sense. Thus the hypothesis (i) of Theorem 1 (below) means: Given $\epsilon > 0$, we can find $\delta = \delta(\epsilon)$ so that $|f(s)/f_0(s) - 1| < \epsilon$ at all points s of D for which $0 < |s| < \delta$.

2. A Tauberian theorem

Let there be given two functions $\varphi(s)$ and $\chi(s)$, and a domain D of the s-plane, satisfying the following conditions:

a) *$\varphi(s)$ and $\chi(s)$ are regular in D, and real and positive on a segment $(0, h]$ of the positive real axis lying in D;*

b) $-\sigma\varphi'(\sigma) \nearrow \infty$ as $\sigma \searrow 0$

[whence $\{\sigma\varphi'(\sigma)\}' \geqq 0$ and therefore

$$\sigma\varphi''(\sigma) \geqq -\varphi'(\sigma) > 0 \tag{3}$$

for sufficiently small positive σ];

c) $$\frac{\delta(\sigma)}{\sigma} \succ \frac{\{\varphi''(\sigma)\}^{\frac{1}{2}}}{-\varphi'(\sigma)} \quad \text{as} \quad \sigma \searrow 0,$$

[2] In a more recent paper, *Neuer Beweis eines Satzes von G. Hardy und S. Ramanujan über das asymptotische Verhalten der Zerfällungskoeffizienten*, Am. Jour. of Math., 62 (1940), pp. 877–880, Avakumović applies substantially the same methods directly to $p(n)$ without explicit formulation of the relevant Tauberian theorem.

where $\delta(\sigma)$ *is the distance of the point* σ *from the complement of* D [so that

(4) $$\delta(\sigma) \leqq \sigma, \qquad (0 < \sigma \leqq h),$$

since the origin cannot belong to D by a) and b)];

d) $$\varphi''(\sigma + z) = O\{\varphi''(\sigma)\},$$
$$\chi(\sigma + z) = O\{\chi(\sigma)\},$$

uniformly for $|z| < \delta(\sigma)$ *when* $\sigma \searrow 0$.

Define:

$$f_0(s) = \chi(s)e^{\varphi(s)}, \qquad F_0(s) = f_0(s)/s,$$
$$A_0(\omega) = \frac{\chi(\sigma)e^{\varphi(\sigma)+\omega\sigma}}{\{2\pi\sigma^2\varphi''(\sigma)\}^{\frac{1}{2}}} = \frac{F_0(\sigma)e^{\omega\sigma}}{\{2\pi\varphi''(\sigma)\}^{\frac{1}{2}}},$$

where $\sigma = \sigma_\omega$ is the solution [existent and unique for sufficiently large ω, by a) and b)] of

(5) $$-\varphi'(\sigma) = \omega, \qquad (0 < \sigma \leqq h).$$

In these circumstances we have

THEOREM 1. *Suppose that the integral* (2) *is convergent for* $\sigma > 0$, *and that*

(i) $f(s) \sim f_0(s)$ *when* $s \to 0$ *in* D;

(ii) $f(s) = O\{f_0(|s|)\}$ *when* $s \to 0$ *in some fixed angle* 'Δ' *of the form* $|t| \leqq \Delta\sigma$, $(0 < \Delta < \infty)$;

(iii) $A(u)$ *is increasing* (*in the wide sense*) *for* $u \geqq 0$.

Then

(6) $$\mathfrak{p}(\Delta) \leqq \overline{\lim_{u\to\infty}} \frac{A(u)}{A_0(u)} \leqq \mathfrak{P}(\Delta),$$

where $\mathfrak{p}(\Delta)$ *and* $\mathfrak{P}(\Delta)$ *depend only on* Δ, *are strictly positive and finite for any given* Δ, $(0 < \Delta < \infty)$, *and tend to* 1 *when* $\Delta \to \infty$.

If (ii) *holds for every fixed* Δ, *then*

(7) $$A(u) \sim A_0(u) \quad as \quad u \to \infty.$$

The last part is obtained from (6) by taking Δ arbitrarily large, and we therefore concentrate on (6). We begin with two lemmas concerning $f_0(s)$.

LEMMA 1. *We have, uniformly for* $|z| \leqq \tau(\sigma) \prec \delta(\sigma)$, *when* $\sigma \searrow 0$,

$$\frac{f_0(\sigma + z)}{f_0(\sigma)} = e^{z\varphi'(\sigma)+\frac{1}{2}z^2\varphi''(\sigma)\{1+o(1)\}+o(1)}$$
$$= e^{z\varphi'(\sigma)+\frac{1}{2}z^2\varphi''(\sigma)}\{1 + o(|z|^2\varphi''(\sigma) + 1)e^{\epsilon|z|^2\varphi''(\sigma)}\}$$

for any fixed $\epsilon > 0$.

By the definition of $f_0(s)$,

(8) $$\frac{f_0(\sigma + z)}{f_0(\sigma)} = \frac{\chi(\sigma + z)}{\chi(\sigma)} e^{\varphi(\sigma+z)-\varphi(\sigma)}.$$

By a) and d) we have

$$\chi'(\sigma + z) = \frac{1}{2\pi i}\int_\Gamma \frac{\chi(\sigma+\zeta)}{(\zeta - z)^2}\,d\zeta = O\left(\frac{\chi(\sigma)}{\delta(\sigma)}\right), \tag{9}$$

where Γ is the circle $|\zeta - z| = \frac{1}{2}\delta(\sigma)$, and all estimates are uniform for $|z| \leqq \tau(\sigma) < \delta(\sigma)$ when $\sigma \searrow 0$; whence by integration

$$\chi(\sigma + z) - \chi(\sigma) = O\left(\frac{\chi(\sigma)\,|z|}{\delta(\sigma)}\right) = O\left(\frac{\chi(\sigma)\tau(\sigma)}{\delta(\sigma)}\right) = o\{\chi(\sigma)\}. \tag{10}$$

Similarly $\varphi''(\sigma + z) - \varphi''(\sigma) = o\{\varphi''(\sigma)\}$, whence by two integrations

$$\varphi(\sigma + z) - \varphi(\sigma) = z\varphi'(\sigma) + \tfrac{1}{2}z^2\varphi''(\sigma) + o\{|z|^2\varphi''(\sigma)\}. \tag{11}$$

Substituting from (10) and (11) into (8) we obtain the first formula of the lemma, and the second follows from it in virtue of the inequality $|e^Z - 1| \leqq |Z|\,e^{|Z|}$.

LEMMA 2. *$f_0(\sigma)$ increases as σ decreases, for sufficiently small positive σ.*

Applying (9) with $z = 0$, we obtain

$$\frac{-f_0'(\sigma)}{f_0(\sigma)} = -\varphi'(\sigma) - \frac{\chi'(\sigma)}{\chi(\sigma)} = -\varphi'(\sigma) + O(1/\delta(\sigma)) \sim -\varphi'(\sigma) \tag{12}$$

as $\sigma \searrow 0$, since, by c), (3) and b),

$$-\delta(\sigma)\varphi'(\sigma) > \{\sigma^2\varphi''(\sigma)\}^{\frac{1}{2}} \geqq \{-\sigma\varphi'(\sigma)\}^{\frac{1}{2}} > 1.$$

Thus $-f_0'(\sigma) > 0$ ultimately. This proves the lemma, but we carry the argument a little further for future reference.

Given ϵ, $(0 < \epsilon < 1)$, we have, by (12) and b),

$$\frac{1}{\epsilon\sigma} < -(1-\epsilon)\varphi'(\sigma) < -\frac{f_0'(\sigma)}{f_0(\sigma)} < -(1+\epsilon)\varphi'(\sigma), \qquad (0 < \sigma \leqq \sigma_\epsilon),$$

whence, integrating over $[\sigma, \sigma_\epsilon]$ and taking exponentials,

$$A_\epsilon\sigma^{-1/\epsilon} < B_\epsilon e^{(1-\epsilon)\varphi(\sigma)} < f_0(\sigma) < C_\epsilon e^{(1+\epsilon)\varphi(\sigma)}, \qquad (0 < \sigma < \sigma_\epsilon), \tag{13}$$

where A_ϵ, B_ϵ, $C_\epsilon > 0$.

PROOF OF THEOREM 1. In the following argument the main variables are σ and ω, connected by (5) and supposed 'sufficiently small' and 'sufficiently large', respectively. Limit operations refer to the process $\sigma \searrow 0$ or the equivalent process $\omega \nearrow \infty$.

In the formula

$$\int_0^\infty A(u)e^{-us}\,du = \frac{f(s)}{s} = F(s), \qquad (\sigma > 0),$$

say, obtained from (2) by partial integration, change s to $\sigma + z$, multiply by $k(z/T)e^{\xi z}\,dz$, where $T > 0$, ξ is real, and $k(z)$ is any function regular along $[0, i]$, and integrate along a path l from 0 to iT. This gives

$$\int_0^\infty A(u)e^{-u\sigma}K(Tu - T\xi)T\,du = \int_0^{iT} F(\sigma + z)k(z/T)e^{\xi z}\,dz, \tag{14}$$

where

$$K(v) = \int_0^i k(Z)e^{-vZ}\,dZ \tag{15}$$

taken along the path L corresponding to l by the transformation $z = TZ$, provided that $\Re(\sigma + z) > 0$ at all points z of l. For L we may take the segment $[0, i]$ or any sufficiently near path giving the same value of $K(v)$.

Take

$$\xi = \xi(\sigma) = \omega + O(\sigma^{-1}) = -\varphi'(\sigma) + O(\sigma^{-1}), \tag{16}$$

$$T = T(\sigma) = \Delta\sigma, \tag{17}$$

with the Δ of hypothesis (ii), and suppose $k(0) \neq 0$. Let $l = l(\sigma)$ be made up of the segment l_1 from 0 to $i\tau$ and the broken line l_2 from $i\tau$ to iT formed of two segments to the left of, and inclined at $\tan^{-1} \tau/\sigma$ to, the segment $[i\tau, iT]$, where $\tau = \tau(\sigma)$ is chosen so that

$$\delta(\sigma) \succ \tau(\sigma) \succ \frac{\sigma\{\varphi''(\sigma)\}^{\frac{1}{2}}}{-\varphi'(\sigma)}. \tag{18}$$

This choice is possible by c); and the path $L = L(\sigma)$ corresponding to $l = l(\sigma)$ tends to coincidence with the segment $[0, i]$, its greatest distance from this segment being $O(\tau/\sigma) = o(1)$ by (18) and (4).

When $z = iy$ is on l_1, the point $\sigma + z$ is in D, by (18), whence, by (16), (17) and hypothesis (i) [since $|z/\sigma| \leqq \tau/\sigma = o(1)$],

$$\frac{F(\sigma + z)k(z/T)e^{\xi z}}{F_0(\sigma)k(0)} = \frac{f(\sigma + z)}{f_0(\sigma)}\cdot\frac{\sigma k(z/\Delta\sigma)}{(\sigma + z)k(0)}\cdot e^{\omega z + O(\tau/\sigma)}$$

$$= \frac{f_0(\sigma + z)}{f_0(\sigma)}\, e^{o(1)}\cdot e^{o(1)}\cdot e^{-z\varphi'(\sigma)+o(1)} = e^{-\frac{1}{2}y^2\varphi''(\sigma)}\{1 + o(y^2\varphi''(\sigma) + 1)e^{\frac{1}{4}y^2\varphi''(\sigma)}\}$$

uniformly in y, by Lemma 1 (with $\epsilon = \frac{1}{4}$). Inserting this in (14) and making the substitution $y\{\varphi''(\sigma)\}^{\frac{1}{2}} = w$, we deduce that the contribution of l_1 to the right hand side of (14) is

$$\frac{ik(0)F_0(\sigma)}{\{\varphi''(\sigma)\}^{\frac{1}{2}}}\left(\int_0^{\eta(\sigma)} e^{-\frac{1}{2}w^2}\,dw + o\int_0^{\eta(\sigma)} (w^2 + 1)e^{-\frac{1}{4}w^2}\,dw\right) \sim \frac{ik(0)F_0(\sigma)(\pi/2)^{\frac{1}{2}}}{\{\varphi''(\sigma)\}^{\frac{1}{2}}},$$

where $\eta(\sigma) = \tau(\sigma)\{\varphi''(\sigma)\}^{\frac{1}{2}} \succ 1$ by (18) and (3).

When $z = x + iy$ is on l_2, the point $\sigma + z$ is in the angle 'Δ' of hypothesis (ii) and $|\sigma + z| > \sigma$ (as may be seen at once from a figure, the lower segment of l_2 being at right angles to the segment $[-\sigma, i\tau]$); whence, by (ii), (16), (17) and Lemma 2,

$$F(\sigma + z)k(z/T)e^{\xi z} = O\left(\frac{f_0(|\sigma + z|)}{|\sigma + z|}\cdot 1\cdot e^{\omega x + O(T/\sigma)}\right) = O\{F_0(\sigma)e^{\omega x}\}$$

uniformly in z. Also $x \leqq 0$ and

$$\left|\frac{dz}{dx}\right| = |1 \mp (\sigma/\tau)i| = O(\sigma/\tau)$$

on each of the segments of l_2. Hence the contribution of l_2 to the right hand side of (14) is

$$O\left(F_0(\sigma)\int_{-\infty}^{0} e^{\omega x}(\sigma/\tau)\,dx\right) = O\left(\frac{F_0(\sigma)\sigma}{\omega\tau}\right) = o\left(\frac{F_0(\sigma)}{\{\varphi''(\sigma)\}^{\frac{1}{2}}}\right),$$

since $\omega\tau = -\varphi'(\sigma)\tau(\sigma) > \sigma\{\varphi''(\sigma)\}^{\frac{1}{2}}$ by (18). Thus

$$\int_0^{iT} F(\sigma + z)k(z/T)e^{\xi z}\,dz \sim \frac{ik(0)(\pi/2)^{\frac{1}{2}}F_0(\sigma)}{\{\varphi''(\sigma)\}^{\frac{1}{2}}} = ik(0)\pi A_0(\omega)e^{-\omega\sigma},$$

by the definition of $A_0(\omega)$. Substituting in (14) and putting $u = \xi + v/T$, we obtain

$$\int_{-T\xi}^{\infty} A(\xi + v/T)e^{-v\sigma/T}K(v)\,dv \sim ik(0)\pi A_0(\omega)e^{(\xi-\omega)\sigma}, \tag{19}$$

in which we note that, by (16), (17) and b),

$$T\xi = -\Delta\sigma\varphi'(\sigma) + O(1) > 1. \tag{20}$$

Take, firstly

$$k(Z) = 2(Z - i).$$

Then

$$K(v) = -\frac{2i}{v} + \frac{2(1 - e^{-iv})}{v^2}, \qquad \Re K(v) = \left(\frac{\sin v/2}{v/2}\right)^2 \geqq 0.$$

Taking real parts in (19), and using (20) and hypothesis (iii) [which implies $A(u) \geqq A(0) = 0$, $(u \geqq 0)$], we deduce that

$$A(\xi - \lambda/T)e^{-\lambda\sigma/T}\int_{-\lambda}^{\lambda}\left(\frac{\sin v/2}{v/2}\right)^2 dv < \{1 + o(1)\}2\pi A_0(\omega)e^{(\xi-\omega)\sigma},$$

where λ is any fixed positive number. Taking

$$\xi = \omega + \lambda/T = \omega + \lambda/\Delta\sigma$$

[an admissible choice by (16)], we infer that

$$\overline{\lim}\,\frac{A(\omega)}{A_0(\omega)} \leqq \frac{2\pi e^{2\lambda/\Delta}}{\int_{-\lambda}^{\lambda}\left(\frac{\sin v/2}{v/2}\right)^2 dv},$$

and the second part of (6) follows on putting $\lambda = \Delta^{\frac{1}{2}}$.

Take, secondly,

$$k(Z) = 2(Z - i) - \frac{e^{\mu Z} - e^{-\mu Z}}{\mu}, \qquad \mu > 0, \qquad \mu \equiv 0 (\text{mod } 2\pi).$$

Then

$$K(v) = -\frac{2i}{v} + (1 - e^{-iv})\left(\frac{2}{v^2} - \frac{1}{\mu(v - \mu)} + \frac{1}{\mu(v + \mu)}\right),$$

$$\Re K(v) = \left(\frac{\sin v/2}{v/2}\right)^2 \frac{\mu^2}{\mu^2 - v^2} \begin{cases} \geqq 0, & (|v| \leqq \mu), \\ \leqq 0, & (|v| \geqq \mu), \end{cases}$$

whence, by a similar argument, with $\xi = \omega - \mu/T$

$$\underline{\lim}\, \frac{A(\omega)}{A_0(\omega)} \geqq \frac{2\pi e^{-2\mu/\Delta}}{\int_{-\mu}^{\mu} \left(\frac{\sin v/2}{v/2}\right)^2 \frac{\mu^2}{\mu^2 - v^2}\, dv}.$$

The denominator here is

$$\int_{-\mu}^{\mu} 2 \sin^2 v/2 \left(\frac{2}{v^2} + \frac{1}{\mu(\mu - v)} + \frac{1}{\mu(\mu + v)}\right) dv$$

$$= \int_{-\mu}^{\mu} \left(\frac{\sin v/2}{v/2}\right)^2 dv + \frac{4}{\mu} \int_0^{2\mu} \frac{\sin^2 w/2}{w}\, dw,$$

and the first part of (6) follows on taking $\mu = 2\pi([\Delta^{\frac{1}{2}}] + 1)$.

An alternative treatment of the range from $i\tau$ to iT in the above proof is possible under certain conditions, for example on the additional assumption b′) of §4 (Theorem 1′), and, in particular, in the special case to be applied in §3. This consists in taking the segment $[i\tau, iT]$ as path of integration and applying an estimate

$$\frac{F(\sigma + iy)k(iy/T)e^{i\xi v}}{F_0(\sigma)} = O(e^{\frac{1}{2}(|\sigma+iy|-\sigma)\phi'(|\sigma+iy|)})$$

$$= O(e^{\delta\sigma^{-1}y^2\phi'(\sigma)}) = O\left(\frac{\sigma}{-\varphi'(\sigma)y^2}\right), \qquad (\delta > 0)$$

to the integrand. [The first step here depends on (ii) and an application of the mean value theorem to $\log f_0(\sigma)$, followed by (12) and b); but the next step seems to require an additional assumption such as b′).] The procedure places the argument on an essentially 'real variable' footing, and this may be considered desirable with a view to the logical status of Theorem 1 in arithmetical applications. It should be observed, however, that the use of a complex variable technique in the original (general) proof does not imply any essential dependence on complex function theory. Thus the fact that the special integrals $K(v)$ are independent of the path $L = L(\sigma)$ is verified by the calculations without reference to Cauchy's general theorem. And, of course, the incidental appeals to Cauchy's theorem in Lemmas 1 and 2 can be avoided in particular applications, like that of §3, by an independent verification of the results.

Special Cases. Take first

$$\varphi(s) = \beta^{-1}(M/s)^{\beta}, \qquad \chi(s) = C(M/s)^{m\beta - \frac{1}{2}}, \qquad (\beta, M, C > 0, m \text{ real}).$$

Then

$$f_0(s) = C(M/s)^{m\beta - \frac{1}{2}} e^{\beta^{-1}(M/s)^{\beta}}, \tag{21}$$

$$A_0(\omega) = \left(\frac{1-\alpha}{2\pi}\right)^{\frac{1}{2}} C(M\omega)^{m\alpha - \frac{1}{2}} e^{\alpha^{-1}(M\omega)^{\alpha}}, \qquad (\alpha = \beta/(\beta + 1)), \tag{22}$$

as may be verified by a simple calculation. For D we may take, for example, any domain of the form $|t| < \Delta_0\sigma^\gamma$ with $\Delta_0 > 0$, $\gamma \geqq 1$ [to satisfy d)], and $\gamma < 1 + \beta/2$ [to satisfy c)]. The actual result to be used in §3, and quoted with special values of M, m, C as (A_2) in §1, is obtained directly by taking $\gamma = 1$ and Δ_0 arbitrarily large; but the hypotheses (i) and (ii) on $f(s)$ become more general as γ is increased (within the permitted range $1 \leqq \gamma < 1 + \beta/2$). Actually it will appear from §4 (Theorem 1′) that in this example we may replace (i) by the corresponding hypothesis along the positive real axis, retaining the complex variable explicitly only in the hypothesis (ii).

The scope of the conditions a), b), c), d) may be illustrated by two further examples, in which we take $\chi(s) = 1$:

$$\varphi(s) = l_1(1/s)l_n(1/s), \qquad \sigma\{l_n(1/\sigma)\}^{-\frac{1}{2}} \prec \delta(\sigma) \leqq \vartheta\sigma, \qquad (0 < \vartheta < 1);$$
$$\varphi(s) = e_n(1/s), \qquad \sigma\{e_n(1/\sigma)\}^{-\frac{1}{2}} \prec \delta(\sigma) \precsim \sigma^2\{e_1(1/\sigma) \cdots e_{n-1}(1/\sigma)\}^{-1}.$$

Here l_n and e_n denote iterated logarithms and exponentials, and the conditions on $\delta(\sigma)$ correspond to c) and d), respectively. A study of such examples, in conjunction with the inequalities (13), shows that $f_0(\sigma)$ must increase more rapidly than any power of $1/\sigma$ when $\sigma \searrow 0$, but that, subject to conditions of regularity and smoothness, any rate of growth above this limit is permitted, the latitude in the choice of D becoming greater with the rate of growth. The extraneous factor $\chi(s)$ adds nothing essential, since it can in practice be absorbed by adding $\log \chi(s)$ to $\varphi(s)$; it serves merely to simplify the calculation of $A_0(\omega)$ in special cases.

3. Application to partitions

THEOREM 2. *Let*

$$0 < \lambda_1 < \lambda_2 < \cdots$$

be a given sequence with

$$(23) \qquad N(u) = Bu^\beta + R(u), \qquad (B > 0, \beta > 0),$$

where $N(u)$ is the number of λ_ν not exceeding u, and

$$(24) \qquad \int_0^u \frac{R(v)}{v}\,dv = b \log u + c + o(1) \qquad \text{as } u \to \infty.$$

For real l let $p(l)$ be the number of solutions of

$$l = r_1\lambda_1 + r_2\lambda_2 + \cdots$$

in integers $r_\nu \geqq 0$, and $p^*(l)$ the number of solutions with $r_\nu = 0$ or 1, (where the formally infinite sum contains a finite but unrestricted number of non-zero terms). For u real and $h > 0$ let

$$P(u) = \sum_{l<u} p(l), \qquad P^*(u) = \sum_{l<u} p^*(l),$$

$$P_h(u) = \frac{P(u) - P(u-h)}{h}, \qquad P_h^*(u) = \frac{P^*(u) - P^*(u-h)}{h}$$

where each summation is over the (discrete) set of l for which the summand is different from zero.

Then, when $u \to \infty$,

(25) $$P(u) \sim \left(\frac{1-\alpha}{2\pi}\right)^{\frac{1}{2}} e^c M^{-(b+\frac{1}{2})\alpha} u^{(b+\frac{1}{2})(1-\alpha)-\frac{1}{2}} e^{\alpha^{-1}(Mu)^\alpha},$$

(26) $$P^*(u) \sim \left(\frac{1-\alpha}{2\pi}\right)^{\frac{1}{2}} 2^b (M^*u)^{-\alpha/2} e^{\alpha^{-1}(M^*u)^\alpha},$$

where

$$\alpha = \frac{\beta}{\beta+1}, \qquad M = \{B\beta\Gamma(\beta+1)\zeta(\beta+1)\}^{1/\beta}, \qquad M^* = (1-2^{-\beta})^{1/\beta} M.$$

Also

(27) $$P_h(u) \sim \left(\frac{1-\alpha}{2\pi}\right)^{\frac{1}{2}} e^c M^{-(b-\frac{1}{2})\alpha} u^{(b-\frac{1}{2})(1-\alpha)-\frac{1}{2}} e^{\alpha^{-1}(Mu)^\alpha},$$

(28) $$P_h^*(u) \sim \left(\frac{1-\alpha}{2\pi}\right)^{\frac{1}{2}} 2^b M^{*\alpha/2} u^{\frac{1}{2}\alpha-1} e^{\alpha^{-1}(M^*u)^\alpha},$$

if h is a positive constant for which the left hand side is an increasing function of u. This condition is certainly satisfied for $P_h(u)$ if h belongs to the given sequence $\{\lambda_\nu\}$, and for $P_h^*(u)$, if $h, 2h, 2^2h, \cdots$ all belong to $\{\lambda_\nu\}$.

The formulae (26) and (28) for $P^*(u)$ and $P_h^*(u)$ remain valid if (24) is replaced by the weaker assumption

(29) $$\int_0^u R(v)\,dv = bu + o(u), \qquad \textit{as } u \to \infty$$

[derivable from (24) by partial integration].

Consider first $P(u)$ and $P_h(u)$. By (23) and (24)

$$N(u) \leqq 2\int_u^{2u} \frac{N(v)}{v}\,dv = 2\left(\int_u^{2u} \frac{Bv^\beta}{v}\,dv + L(2u) - L(u)\right) = O(u^\beta),$$

where $L(u)$ is the left hand side of (24). This suffices to justify the formal transformations which follow, when $\sigma > 0$. By the definition of $p(l)$,

$$\sum_l p(l)e^{-ls} = \prod_{\nu=1}^{\infty} (1-e^{-\lambda_\nu s})^{-1} = g(s),$$

say. By (23) and (24),

$$\log g(s) = -\sum_{\nu=1}^{\infty} \log(1-e^{-\lambda_\nu s}) = -\int_0^\infty \log(1-e^{-us})\,dN(u)$$

(30) $$= \int_0^\infty \frac{s}{e^{us}-1} N(u)\,du = \int_0^\infty \frac{s}{e^{us}-1}(Bu^\beta\,du + u\,dL(u))$$

$$= \frac{B}{s^\beta}\int_{u=0}^{\infty} \frac{(us)^\beta\,d(us)}{e^{us}-1} - \int_{u=0}^{\infty} L(u)\,d\left(\frac{us}{e^{us}-1}\right)$$

$$(31)\qquad = \frac{B\Gamma(\beta+1)\zeta(\beta+1)}{s^\beta} - \int_0^\infty (b\log u + c)\psi'(us)s\,du - \int_0^\infty S(u)\psi'(us)s\,du,$$

where $\psi(s) = s/(e^s - 1)$, and $S(u)$ is $o(1)$ when $u \to \infty$, $O(|\log u|)$ when $u \searrow 0$, and continuous for $u > 0$. The first integral in (31) is

$$\left[(b\log u + c)\frac{us}{e^{us}-1} - b\log(1 - e^{-us})\right]_{u=0}^{\infty} = b\log s - c.$$

And, since $e^{s/2}\psi'(s)$ and $|s|/\sigma$ are bounded in any fixed angle 'Δ' of the type $|t| \leqq \Delta\sigma$, $(0 < \Delta < \infty)$, the second integral is

$$O\left(\int_0^{1/\sqrt{\sigma}} (|\log u| + 1)\sigma\,du\right) + o\left(\int_{1/\sqrt{\sigma}}^\infty e^{-u\sigma/2}\sigma\,du\right) = o(1)$$

when $s \to 0$ in any 'Δ'. Hence

$$\log g(s) = \beta^{-1}(M/s)^\beta - b\log s + c + o(1),$$

$$(32)\qquad g(s) \sim e^c s^{-b} e^{\beta^{-1}(M/s)^\beta},$$

when $s \to 0$ in any fixed 'Δ', where $M^\beta = B\beta\Gamma(\beta+1)\zeta(\beta+1)$.

But

$$\int_0^\infty e^{-us}\,dP(u) = g(s),$$

$$(33)\qquad \int_0^\infty e^{-us}\,dP_h(u) = \frac{1}{h}\left(\int_0^\infty e^{-us}\,dP(u) - \int_0^\infty e^{-(u+h)s}\,dP(u)\right) = \frac{1 - e^{-hs}}{h}\,g(s),$$

and $(1 - e^{-hs})/h \sim s$ when $s \to 0$ in any way. The results (25) and (27) may now be obtained at once by taking $m = (b \pm \frac{1}{2})/\beta$, $C = M^{\frac{1}{2}-m\beta}e^c$, in the first of the special cases of Theorem 1 discussed at the end of §2, since $P(u)$ obviously increases with u and $P_h(u)$ is explicitly assumed to increase.

Finally, the identity (33), which is equivalent to an identity between Dirichlet's series, shows that the step-function $P_h(u)$ increases with u if and only if the series for $(1 - e^{-hs})g(s)$ has positive coefficients; and this is certainly so if h belongs to $\{\lambda_\nu\}$, since then

$$(1 - e^{-hs})g(s) = \prod_{\lambda_\nu \neq h} (1 - e^{-\lambda_\nu s})^{-1}.$$

The formulae for $P^*(u)$ and $P_h^*(u)$ follow similarly from

$$\sum_l p^*(l)e^{-ls} = \prod_{\nu=1}^\infty (1 + e^{-\lambda_\nu s}) = g^*(s),$$

$$g^*(s) \sim 2^b e^{\beta^{-1}(M^*/s)^\beta}, \qquad \text{when } s \to 0 \text{ in any '}\Delta\text{'},$$

where $M^{*\beta} = (1 - 2^{-\beta})M^\beta$. The last relation may be deduced from (32), since $g^*(s) = g(s)/g(2s)$. But the direct proof is more elementary and demands only (29) in place of (24), because $\log(1 + e^{-s})$ and $1/(e^s + 1)$ are regular at $s = 0$,

so that in the formula corresponding to (30) we can use $R(u)\,du$ directly without replacing it by $u\,dL(u)$. The step-function $P_h^*(u)$ increases with u if and only if the Dirichlet's series for $(1 - e^{-hs})g^*(s)$ has positive coefficients; and this is certainly so if the numbers $2^\mu h$, $(\mu = 0, 1, 2, \cdots)$, all belong to $\{\lambda_\nu\}$, since then

$$(1 - e^{-hs})g^*(s) = (1 - e^{-hs}) \prod_{\mu=0}^{\infty} (1 + e^{-2^\mu hs}) \prod_{\lambda_\nu \neq 2^\mu h} (1 + e^{-\lambda_\nu s}) = \prod_{\lambda_\nu \neq 2^\mu h} (1 + e^{-\lambda_\nu s}).$$

Special Cases. The formulae (1) may be deduced at once form (27) and (28), respectively, by taking $\lambda_\nu = \nu$, $(\nu = 1, 2, \cdots)$ and $h = 1$, since

$$(34) \qquad \int_0^u \frac{[v] - v}{v}\,dv = \sum_{\nu \leq u} \log \frac{u}{\nu} - u = -\tfrac{1}{2}\log u - \tfrac{1}{2}\log 2\pi + o(1)$$

as $u \to \infty$ (by Stirling's theorem for $\log n!$), and $\Gamma(2)\zeta(2) = \pi^2/6$. In the case of $q(n)$ the simpler result

$$\int_0^u ([v] - v)\,dv = -\tfrac{1}{2}u + o(u)$$

may be used in place of (34), in virtue of the concluding remark of Theorem 2. The formula for $q(n)$ may, alternatively, be deduced from (27) by taking $\lambda_\nu = 2\nu - 1$, $h = 1$.

An asymptotic formula for the number of partitions of a positive integer n into positive integral k-th powers, where k is any given positive integer, may be deduced similarly from (27) by taking $\lambda_\nu = \nu^k$ and $h = 1$, the verification of (23) and (24), with $\beta = 1/k$, $B = 1$, $b = -\frac{1}{2}$, $c = -\frac{1}{2}k \log 2\pi$ depending on a formula equivalent to (34). This particular problem has been studied in detail by Wright [9], who obtains by more special methods an asymptotic expansion in functions of descending order of magnitude.

Some obvious extensions of the case $\lambda_\nu = \nu^k$ fall within the scope of Theorem 2. Thus we may restrict ν to certain arithmetical progressions, and modify a finite portion of the sequence in any way. In applications of (27) or (28) attention must, of course, be paid to the condition that the left hand side increases with u, at any rate for large u (this being obviously sufficient).

Theorem 2 is a very special deduction from Theorem 1 and could easily be extended to cover other partition functions, such as the number of partitions of n into k_1-th and k_2-th powers.

4. Notes on the conditions of Theorem 1

1) Theorem 1′. *If, in addition to the conditions* a), b), c), d), *we have*

b′) $-\sigma^k\varphi'(\sigma) \searrow 0$ *as* $\sigma \searrow 0$, *for some fixed* k, *then the hypothesis* (i) *of Theorem* 1 *may be replaced by*

(i′) $f(s) \sim f_0(s)$ *when* $s \to 0$ *by real positive values.*

We shall prove the following facts:

(α) On the assumptions a), b), c), d) only, (i)′ and (iii) [or, alternatively,

(i)′ and (ii)] imply (i) with D replaced by a certain sub-domain D_f (depending on f) for which

$$(35)\qquad \delta_f(\sigma) \succ \{\varphi''(\sigma)\}^{-\frac{1}{2}}, \qquad \text{as } \sigma \searrow 0,$$

where $\delta_f(\sigma)$ has the same meaning for D_f as $\delta(\sigma)$ for D;

(β) On the additional assumption b′), the domain D_f is 'admissible' in the sense of conditions a), c), d).

An appeal to Theorem 1 with D replaced by D_f will then establish Theorem 1′.

Choose $\rho(\sigma)$ so that (when $\sigma \searrow 0$)

$$(36)\qquad \{\varphi''(\sigma)\}^{-\frac{1}{2}} \prec \rho(\sigma) \prec \delta(\sigma),$$

as is possible by c) and (3). Consider

$$G(z) = G_\sigma(z) = \frac{f(\sigma + z)}{f_0(\sigma + z)} - 1, \qquad (z = x + iy)$$

in the circle $C(\sigma)$ defined by $|z| \leqq \rho(\sigma)$.

By (iii) and (i)′, $|f(\sigma + z)| \leqq f(\sigma + x) = O\{f_0(\sigma + x)\}$, [or, alternatively, by (ii) and Lemma 2, $f(\sigma + z) = O\{f_0(|\sigma + z|)\} = O(f_0(\sigma + x)\}]$, uniformly for $|z| \leqq \rho(\sigma)$ when $\sigma \searrow 0$; whence, by Lemma 1, $G(z)$ is regular and satisfies

$$|G(z)| = O\left(\frac{f_0(\sigma + x)}{|f_0(\sigma + z)|}\right) + O(1)$$
$$= O(e^{x\varphi'(\sigma)+\frac{1}{2}x^2\varphi''(\sigma)}\, e^{-\Re z\varphi'(\sigma)+\frac{1}{2}|z|^2\varphi''(\sigma)}) + O(1) < e^{2\rho^2(\sigma)\varphi''(\sigma)}$$

in $C(\sigma)$ (if σ is sufficiently small).

By (i)′ we have, since $\rho(\sigma) < \frac{1}{2}\delta(\sigma) \leqq \frac{1}{2}\sigma$ ultimately,

$$|G(z)| < \epsilon(\sigma)$$

on the diameter $[-\rho(\sigma), \rho(\sigma)]$, where $\epsilon(\sigma) = \epsilon_f(\sigma) \to 0$ as $\sigma \searrow 0$ and is independent of the choice of $\rho(\sigma)$ in the range (36).

It follows, by a well known convexity theorem [applied separately to the upper and lower halves of $C(\sigma)$], that, in the region bounded by the two circular arcs $-\rho(\sigma)$, $\pm i\lambda\rho(\sigma)$, $\rho(\sigma)$, $(0 < \lambda < 1)$, and so in particular in the circle $|z| < \lambda\rho(\sigma)$,

$$(37)\qquad |G(z)| < \epsilon^{\vartheta}(\sigma)e^{(1-\vartheta)2\rho^2(\sigma)\varphi''(\sigma)},$$

where $(1 - \vartheta)\pi$ is the angle at which the two circular arcs intersect, so that $0 < \vartheta < 1$ and ϑ depends only on λ.

Take a fixed λ and choose $\rho(\sigma) = \rho_f(\sigma)$ in the range (36) so that the right hand side of (37) is $< \epsilon^{\frac{1}{2}\vartheta}(\sigma)$. Then we shall have $f(s) \sim f_0(s)$ when $s \to 0$ in the sub-domain

$$D_f = \sum_{0<\sigma\leqq h_f} d_f(\sigma)$$

of D, where $d_f(\sigma)$ is the circular domain of the s-plane with centre σ and radius

$\lambda\rho_f(\sigma)$, h_f is a sufficiently small positive number, and $\sum$ denotes logical summation over σ. Also $\delta_f(\sigma) \geqq \lambda\rho_f(\sigma) > \{\varphi''(\sigma)\}^{-\frac{1}{2}}$ by (36). This proves (α).

Since D_f lies in D, contains the segment $(0, h_f]$, and satisfies (35), it will satisfy all the conditions imposed on D in a), c), d), if $\sigma\varphi''(\sigma) = O\{-\varphi'(\sigma)\}$. But this is so on the assumption b′), which implies that $\{\sigma^k\varphi'(\sigma)\}' \leqq 0$ and therefore

$$\sigma\varphi''(\sigma) \leqq -k\varphi'(\sigma).$$

This proves (β) and completes the proof of Theorem 1′.

2) In Theorem 1′ we may regard (iii) and (ii) (for every Δ) as Tauberian conditions which convert the generally false inference (i)′ $\to$ (7) into a true proposition. An example showing that (iii) alone is not a sufficient Tauberian condition has been constructed by Avakumovic and Karamata [3] (353, c). In this example (i)′ and (iii) are satisfied but the $\underline{\lim}$ and $\overline{\lim}$ in (6) are 0 and ∞.

Another example proving less in this direction, but illustrating the effect of the supplementary hypothesis (ii), may be constructed as follows. With the special functions (21) and (22), and any fixed $U > 0$, we can show, by the saddle point method, that

$$\int_U^\infty e^{-us}\, dA_0(u) \sim f_0(s), \tag{38}$$

when $s \to 0$ in a certain angle $|t| \leqq \delta\sigma$, $(\delta > 0)$; whence, by a change of the variables and of the path of integration,

$$\int_U^\infty e^{-us}\, dA_0(\rho u) \sim f_0(s/\rho), \tag{39}$$

when $s \to 0$ in $|t| \leqq \delta'\sigma$, $(0 < \delta' < \delta)$, if $\rho = re^{i\vartheta}$ is fixed with $r > 0$, and $\vartheta > 0$ and sufficiently small.

Choose ρ so that $\Re\rho^\alpha = 1$. Then $\Re\rho^\beta < 1$, since

$$(\cos\beta\vartheta)^{1/\beta} < (\cos\alpha\vartheta)^{1/\alpha},$$

as may be verified without difficulty (for small positive ϑ) since $\beta > \alpha > 0$. Take

$$A(u) = A_0(u) + \lambda\Re A_0(\rho u), \quad (u \geqq U), \qquad A(u) = 0, \quad (0 \leqq u < U),$$

with $\lambda > 0$. Then, since $\Re\rho^\alpha = 1$,

$$\overline{\underline{\lim_{u\to\infty}}}\, \frac{A(u)}{A_0(u)} = 1 \pm \lambda', \qquad (\lambda' > 0). \tag{40}$$

so that (7) is false. On the other hand, since $\Re\rho^\beta < 1$, $f_0(\sigma/\rho) = o\{f_0(\sigma)\}$ as $\sigma \searrow 0$, whence, by (38) and (39),

$$f(s) = \int_0^\infty e^{-us}\, dA(u) \sim f_0(s) \tag{41}$$

when $s \to 0$ by real positive values, so that (i)′ is satisfied. Also, (iii) is satisfied if λ is sufficiently small and U sufficiently large, as may be verified by differentiation.

In this example, (41) holds in the angle $|t| \leqq \Delta\sigma$ if Δ is sufficiently small, so that we have (ii) and therefore (6) for some $\Delta > 0$. On the other hand, (40) shows that, apart from the precise forms of $\mathfrak{p}(\Delta)$ and $\mathfrak{P}(\Delta)$, the conclusion (6) cannot be improved, at any rate for the smaller values of Δ.

The actual form of the supplementary hypothesis (ii) was adopted with a view to applications, but it is naturally not the only possible form.

3) The distinction between Theorem 1′ and the Hardy-Littlewood case $f_0(s) = s^{-a}$, $(a > 0)$, in which (iii) alone is a sufficient Tauberian condition for an inference of the form (i)′ → (7), can be explained very simply in terms of the ideas of this paper. A valid (if somewhat sophisticated) proof of the Hardy-Littlewood theorem [quoted in the case $a = 1$ as (A_1) in §1] can be constructed by applying, first, a convexity argument, substantially as in Theorem 1′ (α) [with $\chi(s) = 1$ and $\varphi(s) = a \log 1/s$] or as a Phragmén-Lindelöf theorem, to deduce from (i)′ and (iii) that $f(s) \sim f_0(s)$ when $s \to 0$ in any fixed angle $|t| \leqq \Delta\sigma$, $(0 < \Delta < \infty)$, and, then, a slightly modified form of the argument of Theorem 1, with $\tau = T = \Delta\sigma$ and an arbitrarily large Δ, to deduce (7) with $A_0(u) = u^a/\Gamma(a + 1)$. Thus the convexity argument is completely successful in removing the complex variable from the hypotheses in the Hardy-Littlewood case, but only partially successful in Theorem 1′, the essential difference lying in the fact that $\{\varphi''(\sigma)\}^{-\frac{1}{2}}$ is of the same order as σ in the former case, but of lower order in the latter.

4) If we ask what can be proved about $A(u)$ when we simply omit (ii) from the hypotheses of Theorem 1′, or, more generally, when we replace the hypotheses (i) and (ii) of Theorem 1 by (i)′, a partial answer can be given by the methods of this paper, though the question belongs more naturally to the order of ideas of Martin and Wiener [5]. We can first use the convexity argument of Theorem 1′(α) to extend (i)′ to the region $|t| \leqq C\{\varphi''(\sigma)\}^{-\frac{1}{2}}$, where C is any positive constant, and then take $\tau = T = C\{\varphi''(\sigma)\}^{-\frac{1}{2}}$ in the argument of Theorem 1 [or in the corresponding argument with $A(u)\,du$ replaced by $dA(u)$ and $F(s)$ by $f(s)$]. The general effect of this, when we reject the parts of the range of integration where $\Re K(v)$ is relatively small and use the condition $A(u) \geqq 0$ [or the condition '$dA(u) \geqq 0$'], is to give information about certain weighted averages of $A(u)\,du$ [or of $dA(u)$] over an interval about the point $u = \omega = -\varphi'(\sigma)$ whose length is of order $1/T$, i.e. of order $\{\varphi''(\sigma)\}^{\frac{1}{2}}$. But this is just the kind of result that Martin and Wiener obtain by real variable methods more appropriate to the form of the problem. The actual theorems proved by Martin and Wiener do not emerge readily by the method outlined above, which is somewhat artificial in this context, but the discussion serves at any rate to explain the effect of the various hypotheses concerning $f(s)$. The extension from the real form (i)′ to the complex form '(i) + (ii)' reduces the length of the interval of averaging from $c\{\varphi''(\sigma)\}^{\frac{1}{2}}$ to c/σ, and this enables us to draw the conclusion (6) when $A(u)$

is monotonic, because $e^{-u\sigma}$ varies only by a constant factor $e^{\pm c}$, when u ranges over an interval of length c/σ; and the sharper conclusion (7) is possible if c can be taken arbitrarily small, because $e^{\pm c} \to 1$ when $c \to 0$.

THE INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY

REFERENCES

(1) V. G. Avakumović, *Über das Verhalten Laplacescher Integrale an der Konvergenzgrenze mit neuem Beweis eines Satzes von Hardy-Ramanujan über das asymptotische Verhalten der Zerfällungskoeffizienten*, (2 Congr. Interbalkan, des Math., Bucarest, 12-IX-1937), Bull. Math. Soc. Roum. Sci. **40**, Nr. 1/2 (1938), 101–106.

(2) V. G. Avakumović, *Bemerkungen über Laplacesche Integrale, deren Wachstum von Exponentialcharakter ist.* I *und* II, Math. Zeitschrift, **46** (1940), 62-69.

(3) V. Avakumović and J. Karamata, *Über einige Taubersche Sätze, deren Asymptotik von Exponentialcharakter ist.* I, Math. Zeitschrift, **41** (1936), 345-356.

(4) G. H. Hardy and S. Ramanujan, *Asymptotic formulae in combinatory analysis*, Proc. London Math. Soc. (2), **17** (1918), 75–115 [= Collected works of S. Ramanujan, 276–309 (Paper 36)].

(5) W. T. Martin and N. Wiener, *Taylor's series of functions of smooth growth in the unit circle*, Duke Math. Journal, **4** (1938), 384–392.

(6) H. Rademacher, *On the partition function* $p(n)$, Proc. London Math. Soc. (2), **43** (1937), 241–254.

(7) J. V. Uspensky, *Asymptotic expressions of numerical functions occurring in problems of partition of numbers* [in Russian], Bull. Acad. Sci. de Russie (6), **14** (1920), 199–218.

(8) N. Wiener and H. R. Pitt, *A generalization of Ikehara's theorem*, Journal of Math. and Phys. M.I.T., **17** (1938), 247–258.

(9) E. M. Wright, *Asymptotic partition formulae.* III. *Partitions into* k^{th} *powers*, Acta Math., **63** (1934), 143–191.

ANNALS OF MATHEMATICS
Vol. 42, No. 5, December, 1941

BEITRÄGE ZUR TOPOLOGIE DER GRUPPEN-MANNIGFALTIGKEITEN

VON HANS SAMELSON

(Received October 18, 1940)

EINLEITUNG

Die Theorie der Lie'schen Gruppen war ursprünglich eine "lokale" Theorie: man betrachtete nur eine Umgebung des Einselementes einer Gruppe, das, was man heute einen Gruppenkeim oder eine lokale Gruppe nennt [1].[1] Diese Einschränkung, die dem Gruppenbegriff eigentlich nicht angemessen ist, wurde in der neueren Forschung aufgegeben. Man betrachtet heute die Algebra der *ganzen* Gruppe und die Geometrie im *Grossen* des die Gruppe repräsentierenden Raumes, der "Gruppen-Mannigfaltigkeit". Die damit auftretenden geometrischen Fragen sind im wesentlichen topologische Probleme. Die hierhergehörigen Ergebnisse, die im Sommer 1935 bekannt waren, hat E. Cartan in einem Vortrag "Sur la topologie des groupes de Lie" zusammengestellt [2]. (In der vorliegenden Arbeit werden nur kompakte Gruppen, deren Räume also geschlossene Mannigfaltigkeiten sind, behandelt.)

Der topologische Bau einer Mannigfaltigkeit wird in erster Linie beschrieben durch ihre Betti'schen Gruppen, sowie durch ihren Homologie-Ring, den man aus den Betti'schen Gruppen bekommt, wenn man die Schnitt-Bildung als Multiplikation einführt.[2] Für *Gruppen*-Mannigfaltigkeiten hat dieser Ring eine besondere Bedeutung: er ist—nach Cartan—isomorph dem Ring der invarianten Differentialformen der Gruppe. Daraus ergibt sich weiter, dass das topologische Problem, die Struktur des Homologie-Ringes einer Gruppe zu bestimmen, äquivalent ist mit einem algebraischen Problem, das von den Invarianten linearer Gruppen handelt. Auf beiden Wegen zur Untersuchung des Ringes sind aber die Schwierigkeiten so gross, dass heute noch nicht die Ringe aller bekannten Gruppen ermittelt sind.

Immerhin ist diese Aufgabe gelöst für die Gruppen, die Weyl die "klassischen" nennt—also die orthogonalen, die unitären und die symplektischen—, und zwar von Pontrjagin auf dem topologischen [3], von R. Brauer auf dem algebraischen Wege [4]. Es stellte sich dabei heraus: Die genannten Gruppen haben die gleichen Homologie-Ringe wie gewisse topologische Produkte von Sphären. Die Dimensionszahlen dieser Sphären, die ja dann den Ring bestimmen, lassen sich für jede dieser Gruppen angeben. Für die Poincaré'schen Polynome—also

[1] Die Nummern im Text verweisen auf das Literaturverzeichnis am Schluss der Arbeit.

[2] Wir verwenden als Koeffizientenberich in dieser Arbeit immer den Körper der rationalen Zahlen (oder, was fast auf dasselbe herauskommt, den Körper der reellen Zahlen; dieser Körper ist bei der im nächsten Satz genannten Isomorphie mit dem Ring der invarianten Differentialformen benützt).

für die Polynome in einer Unbestimmten t, deren Koeffizienten die Betti'schen Zahlen der betreffenden Mannigfaltigkeit sind—folgt, dass sie in Faktoren der Gestalt $(1 + t^m)$ zerfallen; dabei sind die Exponenten m die erwähnten Dimensionszahlen.

Diese wichtigen Untersuchungen von Pontrjagin und Brauer—(ebenso wie eine etwas jüngere topologische Arbeit von Ehresmann [5], in der dieselben Ergebnisse erhalten werden)—haben einen "speziellen" Charakter: es wird jedesmal eine spezielle Klasse von Gruppen betrachtet, und es werden jedesmal ganz spezielle Eigenschaften dieser Gruppen benützt. Daher erhebt sich die Frage nach einer allgemeinen topologischen Theorie der Gruppenmannigfaltigkeiten, aus der man die erwähnten Sätze über die klassischen Gruppen durch Spezialisierung gewinnen könnte. So schliesst auch Cartan seinen genannten Vortrag nach einem Bericht über die Pontrjagin-Brauer'schen Untersuchungen mit den Worten: "Il faut espérer qu'on trouvera aussi une raison de portée générale expliquant la forme si particulière des polynomes de Poincaré des groupes simples clos." [6]

In dieser Hinsicht ist nun ein wichtiger Schritt der folgende Satz von Hopf: [7] "Für jede (geschlossene) Gruppe ist der Homologie-Ring isomorph dem Ring eines gewissen topologischen Produktes von Sphären, deren Dimensionen ungerade sind"; es gilt also allgemein das, was Pontrjagin und Brauer für die klassischen Gruppen festgestellt haben—insbesondere zerfällt das Poincaré'sche Polynom *jeder* Gruppe in Faktoren der Gestalt $(1 + t^m)$ mit ungeraden Exponenten m. Dieser Satz besitzt offenbar die gewünschte Allgemeinheit; jedoch wird in seinem Beweis die Gruppen-Eigenschaft nicht vollständig ausgenützt und vor allem ein wichtiges Hilfsmittel nicht herangezogen, das Pontrjagin in seiner Arbeit [3] eingeführt hat: eine Multiplikation von Zyklen mit Hilfe der Gruppenmultiplikation, die kurz so aussieht: Durchläuft der Punkt p den Zyklus x und der Punkt q den Zyklus y, dann durchläuft der Produktpunkt pq den "Produktzyklus" von x und y.

Auf der Betrachtung dieser "Pontrjagin'schen" Multiplikation und ihrer Kombination mit dem Satz von Hopf basiert nun die vorliegende Arbeit. Und zwar beweisen wir in Kap. I, §3 einen Satz (den "Aufspannsatz") über die Struktur des Ringes einer Gruppe, dessen Beweis für die klassischen Gruppen einen wesentlichen Teil des Inhaltes der Pontrjagin'schen Arbeit bildet und dessen Gültigkeit für alle Gruppen von Cartan [8] vermutet wurde. Der Inhalt dieses Satzes ist eine ganz bestimmte Dualität zwischen der Gruppen- (Pontrjagin'schen)-Multiplikation der Zyklen und der Schnitt-Bildung der Zyklen; dadurch wird die Gruppe in eine noch engere Analogie zu einem Sphären-Produkt gebracht als durch den Satz von Hopf; es handelt sich um einen "doppelten" Isomorphismus.

Der §1 von Kap. I enthält eine Zusammenstellung bekannter Tatsachen über Homologie-Ringe von Mannigfaltigkeiten, und die Definition und Eigenschaften der (bei Hopf [7] eingeführten) Begriffe "maximales" und "minimales" Element eines Homologie-Ringes. In §2 wird der Satz von Hopf zitiert, und

es werden die für das folgende grundlegenden Eigenschaften des Ringes einer Gruppe dargestellt. Der §3 enthält den Beweis des oben genannten "Aufspannsatzes".

Auf Grund dieses Satzes werden dann in Kap. II die beiden wichtigsten gruppentheoretischen Begriffe "homomorphe Abbildung" und "Untergruppe" in topologischer Hinsicht untersucht. Es wird die Frage diskutiert, wie die Mannigfaltigkeit einer Untergruppe in die Gruppe eingelagert ist, d.h. welche Zyklen aus der Untergruppe in der Gruppe homolog Null sind; dabei ergeben sich einfache Gesetze. Diese spielen im folgenden eine Rolle.

Eine kontinuierliche Gruppe G ist gewöhnlich gegeben als transitive Transformationsgruppe einer Mannigfaltigkeit W, des "Wirkungsraumes" ("espace homogène") [9]. Diejenigen Transformationen, die einen bestimmten Punkt von W festhalten, bilden eine Untergruppe U von G, die sogenannte "Isotropiegruppe" von W, die bis auf innere Automorphismen von G unabhängig von der Wahl des festen Punktes ist. Die Nebengruppen, in welche G nach U zerfällt, stehen in eineindeutiger und stetiger Beziehung zu den Punkten von W; geometrisch gesehen liegt eine "Faserung" von G in die (untereinander homöomorphen) Nebengruppen von U vor [10].

Eines der Hauptprobleme der weiteren Theorie besteht nun in der Untersuchung der Beziehungen zwischen G, U und W. Einerseits ist das der natürliche Ansatzpunkt zur Untersuchung einer als Transformationsgruppe gegebenen Gruppe; andererseits sind die Wirkungsräume, also die Mannigfaltigkeiten, die transitive Transformationsgruppen zulassen, an und für sich interessant. Hier hat man nun Möglichkeiten, wenn man die Sätze der Kap. I und II dieser Arbeit verknüpft mit topologischen Sätzen über die Faserungen von Mannigfaltigkeiten. Dazu werden in Kap. III und IV einige Beiträge geliefert.

Das Kap. III behandelt den Fall, dass der Wirkungsraum W eine Sphäre ist. Das Ergebnis ist, dass man auf eine sehr einfache Weise aus dem Homologie-Ring der Isotropiegruppe U und der Dimension der Wirkungs-Sphäre W den Ring der Gruppe G bestimmen kann. Dieser allgemeine Satz gibt mühelos durch eine Rekursion die Pontrjagin-Brauer'schen Resultate über die Ringe der "klassischen" Gruppen; denn diese Gruppen sind bekanntlich transitive Transformationsgruppen von Sphären, und die zugehörigen Isotropiegruppen sind auch klassische Gruppen.

Im Kap. IV werden die Beziehungen zwischen G, U und W ohne spezielle Voraussetzungen über W untersucht. Unter anderem wird der Satz bewiesen: Ist U nicht homolog 0 in G, dann ist der Homologie-Ring von G isomorph dem Ring des topologischen Produktes von U und W. Dabei ergibt sich noch: Ein Wirkungsraum W, dessen Isotropiegruppe U nicht homolog 0 in der Gruppe G ist, hat den gleichen Ring wie ein gewisses Produkt von Sphären ungerader Dimensionen. Das ist eine Verallgemeinerung des Satzes von Hopf über den Ring einer Gruppen-Mannigfaltigkeit (eine Gruppe lässt sich auffassen als Wirkungsraum, dessen Isotropiegruppe ein Punkt ist, und ein Punkt ist nie homolog 0).

KAPITEL I

EIN SATZ ÜBER GRUPPEN-MANNIGFALTIGKEITEN

Die §§1 und 2 enthalten eine geeignete Zusammenstellung bekannter Dinge (im wesentlichen ohne Beweise) und einen Bericht über eine Arbeit von Hopf über Gruppenmannigfaltigkeiten, auf die sich die vorliegende Arbeit stützt. In §3 wird ein allgemeiner Satz über Gruppenmannigfaltigkeiten bewiesen.

1. Der Homologie-Ring einer Mannigfaltigkeit. Irreduzible Erzeugendensysteme, maximale und minimale Elemente

1. M sei eine geschlossene, orientierbare Mannigfaltigkeit der Dimension n. Die Homologieklassen von M, die Elemente der Betti'schen Gruppe $\mathfrak{B}(M)$, bilden einen Ring, den Homologie- oder Schnittring $\mathfrak{R}(M)$, wenn man die Addition durch die Addition in der Betti'schen Gruppe und das Produkt durch die Schnittbildung erklärt [11]. *Dabei sollen in dieser Arbeit immer die rationalen Zahlen als Koeffizienten dienen.*

$\mathfrak{R}(M)$ enthält ein Eins-Element, nämlich den orientierten n-dimensionalen Grundzyklus von M; es wird mit 1 bezeichnet. Die Dimension eines homogendimensionalen Elementes z von $\mathfrak{R}(M)$ bezeichnen wir mit $d(z)$; die Zahl $\delta(z) = n - d(z)$ heisse die duale Dimension von z. Das Null-Element 0 hat jede Dimension.

Für homogen-dimensionale Elemente z, z' ist auch das Produkt oder der Schnitt $z \cdot z'$ homogen-dimensional und zwar gilt:

$$\delta(z \cdot z') = \delta(z) + \delta(z').$$

Weiter ist:

$$z \cdot z' = (-1)^{\delta(z)\cdot\delta(z')} z' \cdot z,$$

also speziell:

$$z \cdot z = 0 \qquad \text{bei ungeradem } \delta(z).$$

Die rationalen Vielfachen der 1 heissen die Skalare.

Wo im folgenden Homologieklassen auftreten, sollen sie immer, wenn nicht anders bemerkt, homogen-dimensional sein.

2. $\mathfrak{B}_r(M)$, oder kurz $\mathfrak{B}_r$, sei die r-te Betti'sche Gruppe von M, also die additive Gruppe der r-dimensionalen Elemente von $\mathfrak{R}(M)$; ihr Rang p_r ist die r-te Betti'sche Zahl von M. Die volle Betti'sche Gruppe $\mathfrak{B}(M)$ ist die direkte Summe

$$\mathfrak{B}_0(M) + \mathfrak{B}_1(M) + \cdots + \mathfrak{B}_n(M).$$

Für $r < n$ sei $\mathfrak{U}_r$ die Gruppe derjenigen Elemente von $\mathfrak{B}_r$, die sich aus den Elementen der Gruppen

$$\mathfrak{B}_{n-1}, \mathfrak{B}_{n-2}, \cdots, \mathfrak{B}_{r+1}$$

durch Multiplikation und Addition erzeugen lassen (vgl. [7], Nr. 30), die sich also als $\sum x_h \cdot y_h$ mit $\delta(x_h) > 0$ und $\delta(y_h) > 0$ schreiben lassen; die Elemente von $\mathfrak{U}_r$ heissen "zusammengesetzt."

Ein nicht-skalares Element heisst maximal, wenn es nicht zusammengesetzt ist, wenn es sich also nicht als Summe von Produkten von höherdimensionalen Elementen darstellen lässt. Ist z maximal und w zusammengesetzt, dann ist auch $z + w$ maximal, weil $\mathfrak{U}_r$ eine Gruppe ist. Die Skalare werden nicht als maximal angesehen; die höchstdimensionalen unter den nicht-skalaren, nicht-verschwindenden Elementen sind immer maximal.

Man stelle $\mathfrak{B}_r$ als direkte Summe $\mathfrak{U}_r + \mathfrak{Z}_r$ dar; das ist möglich, weil die rationalen Zahlen als Koeffizienten dienen. Die Elemente von $\mathfrak{Z}_r$ sind dann maximal. Wählt man in jeder Gruppe $\mathfrak{Z}_r$ eine Basis, so bildet die Gesamtheit dieser Basiselemente mit der 1 zusammen, wie man leicht sieht, ein Erzeugendensystem des Ringes $\mathfrak{R}(M)$; d.h.: jedes Element von $\mathfrak{R}(M)$ lässt sich als Polynom in den Elementen dieses Systems schreiben. Man bestätigt auch leicht, dass ein irreduzibles Erzeugendensystem vorliegt, d.h. dass kein echtes Teilsystem ein Erzeugendensystem ist.

Der Rang der Gruppe $\mathfrak{Z}_{n-r}$, also auch der Rang der Restklassengruppe $\mathfrak{B}_{n-r} - \mathfrak{U}_{n-r}$, werde l_r genannt. (Rang = Maximalzahl linear unabhängiger Elemente.) Man stellt fest, dass die nicht-skalaren Elemente eines beliebigen irreduziblen Erzeugendensystems $(1, z_1, z_2, \cdots, z_l)$ von $\mathfrak{R}(M)$ maximal sind, und dass die Zahl derjenigen z_i, für die $\delta(z_i) = r$ ist, immer gleich l_r ist; die Anzahl der Elemente eines solchen Systems (ohne die 1) ist also immer

$$l = l_1 + l_2 + \cdots + l_n . \qquad \text{([7], Nr. 31)}$$

Ersetzt man die Elemente z_i eines solchen Systems durch $z_i + y_i$, wo die Elemente y_i nicht maximal sind, so erhält man wieder ein irreduzibles Erzeugendensystem; das kommt daher, dass die z_i Repräsentanten für die Restklassengruppe $\mathfrak{B}_r - \mathfrak{U}_r$ sind.

3. Unter den Vielfachen $x \cdot v$ eines Elementes v von $\mathfrak{R}(M)$, wo x alle Elemente von $\mathfrak{R}(M)$ durchläuft, treten immer auf:

a) die rationalen Vielfachen von v, deren Dimension gleich der von v ist—man setze für x einen beliebigen Skalar ein—und

b) alle null-dimensionalen Elemente, d.h. der Punkt mit einem beliebigen Koeffizienten, wenn $v \neq 0$ ist—wegen des Poincaré-Veblen'schen Dualitätssatzes.

Wir definieren nun:

Ein Element v von $\mathfrak{R}(M)$ heisst minimal, wenn es keine anderen Vielfachen besitzt, d. h. wenn aus $0 < d(x \cdot v) < d(v)$ *folgt:* $x \cdot v = 0$. (Statt $0 < d(x \cdot v) < d(v)$ kann man auch sagen: $0 < \delta(x) < d(v)$.) (Vgl. [7], Nr. 32)

Die nulldimensionalen Elemente gelten nicht als minimal, so wie die n-dimensionalen Elemente nicht als maximal gelten. Dagegen ist die Null minimal.

Es gilt der folgende Satz:

Das s-dimensionale Element v ist dann und nur dann minimal, wenn es ein

Annulator der Gruppe $\mathfrak{U}_{n-s}$ ist, d.h. wenn $u \cdot v = 0$ ist für jedes Element u von $\mathfrak{U}_{n-s}$ ($s = 1, \cdots, n$) ([7], Nr. 33). Daraus entnimmt man: die s-dimensionalen minimalen Elemente bilden eine additive Gruppe $\mathfrak{V}_s$, eine Untergruppe von $\mathfrak{B}_s$. Aus dem Poincaré-Veblen'schen Dualitätssatz folgt leicht, dass der Rang von $\mathfrak{V}_s$ gleich l_s ist, nämlich gleich dem Rang von $\mathfrak{Z}_{n-s}$ in der direkten Summe $\mathfrak{B}_{n-s} = \mathfrak{U}_{n-s} + \mathfrak{Z}_{n-s}$; $\mathfrak{V}_s$ und $\mathfrak{Z}_{n-s}$ sind gewissermassen dual zueinander. Die direkte Summe $\mathfrak{V}_1 + \mathfrak{V}_2 + \cdots + \mathfrak{V}_n$ aller dieser Gruppen heisst die Gruppe der minimalen Elemente von $\mathfrak{R}(M)$; sie wird mit $\mathfrak{V}(M)$, oder kurz mit $\mathfrak{V}$ bezeichnet. Ihr Rang ist

$$l = l_1 + l_2 + \cdots + l_n .$$

4. M und M' seien zwei Mannigfaltigkeiten, f eine Abbildung von M in M'.[3] f bildet den Ring $\mathfrak{R}(M)$ in den Ring $\mathfrak{R}(M')$ ab. Dann gilt der wichtige Invarianzsatz:

Ist v ein minimales Element von $\mathfrak{R}(M)$, so ist sein Bild $f(v)$ ein minimales Element von $\mathfrak{R}(M')$. ([7], Nr. 34)

Eine spezielle Folgerung hieraus ist:

Ein Element von $\mathfrak{R}(M)$, das durch das stetige Bild einer Sphäre S_n ($n > 0$) repräsentiert wird, ist ein minimales Element. Denn in $\mathfrak{R}(S_n)$ ist der Grundzyklus S_n—die 1 von $\mathfrak{R}(S_n)$—minimal.

5. M und M' seien zwei geschlossene, orientierbare Mannigfaltigkeiten; f sei eine Abbildung von M in M'. Durch f wird eine Abbildung von $\mathfrak{R}(M)$ in $\mathfrak{R}(M')$ bewirkt, die auch f heisse; sie ist additiv, aber im allgemeinen nicht multiplikativ homomorph. Dann existiert eine Abbildung φ von $\mathfrak{R}(M')$ in $\mathfrak{R}(M)$—der Umkehrungshomomorphismus—mit folgenden Eigenschaften: [12]

1) φ ist ein Ring-Homomorphismus, also additiv und multiplikativ homomorph,

2) φ ist mit f durch die Funktionalgleichung

$$f(\varphi(z') \cdot x) = z' \cdot f(x)$$

verknüpft, in der x ein beliebiges Element von $\mathfrak{R}(M)$ und z' ein beliebiges Element von $\mathfrak{R}(M')$ ist,

3) ist z' ein homogen-dimensionales Element von $\mathfrak{R}(M')$, so ist auch $\varphi(z')$ homogen-dimensional, und zwar ist

$$\delta(\varphi(z')) = \delta(z'),$$

anders geschrieben:

$$d(\varphi(z')) = d(z') + d(M) - d(M').$$

(Vgl. [7], Nr. 25) Mit Hilfe dieser Eigenschaften lässt sich allgemein bei gegebener Abbildung f, bzw. φ, die Abbildung φ, bzw. f, anschreiben; wir werden

[3] Alle in dieser Arbeit vorkommenden Abbildungen sind eindeutig und stetig.

jedoch diese Beziehung, die wir nur in spezieller Weise brauchen, jedesmal direkt ableiten.

6. Es soll jetzt noch ein Hilfssatz abgeleitet werden, der später gebraucht wird. Es sei z_0, bzw. z_0', der einfach gezählte Punkt von M, bzw. M'.

HILFSSATZ: *Ist* $\varphi(z_0') \neq 0$, *so ist* f *eine Abbildung von* $\mathfrak{R}(M)$ ***auf*** $\mathfrak{R}(M')$.

Es sei also $\varphi(z_0') \neq 0$; wir zeigen zunächst: Ist x' von Null verschieden, dann ist auch $\varphi(x')$ nicht Null. Weil x' nicht Null ist, gibt es ein duales y' mit $y' \cdot x' = z_0'$. Wegen der Multiplikativität von φ ist $\varphi(y') \cdot \varphi(x') = \varphi(y' \cdot x') = \varphi(z_0') \neq 0$. Also ist auch $\varphi(x') \neq 0$. Die Gruppe derjenigen Elemente von $\mathfrak{B}(M')$, die bei der Abbildung f als Bilder von Elementen von $\mathfrak{B}(M)$ auftreten, die Bildgruppe $f(\mathfrak{B}(M))$, hat dann die folgende Eigenschaft:

Sie enthält zu einem beliebigen Element $x'(\neq 0)$ von $\mathfrak{B}(M')$ ein Element y' mit $x' \cdot y' = z_0'$.

Denn aus $x' \neq 0$ folgt $\varphi(x') \neq 0$; also gibt es ein y so, dass $\varphi(x') \cdot y = z_0$ ist; übt man auf diese Gleichung f aus und beachtet die Funktionalgleichung für φ, so erhält man $x' \cdot f(y) = z_0'$. Die Behauptung ist also mit $y' = f(y)$ erfüllt.

Aber aus dieser Eigenschaft der Bildgruppe folgt nach dem Poincaré-Veblen'schen Dualitätssatz sofort, dass die Bildgruppe mit $\mathfrak{B}(M')$ zusammenfällt, was zu beweisen war.

7. Zu zwei Mannigfaltigkeiten M und M' gehört die Produktmannigfaltigkeit $M \times M'$. Ueber die Homologieeigenschaften von $M \times M'$ ist bekannt [13] (vgl. [7], Nr. 19):

Zu jedem Element x von $\mathfrak{R}(M)$ und jedem Element x' von $\mathfrak{R}(M')$ gehört ein Element $x \times x'$ von $\mathfrak{R}(M \times M')$; diese Produktbildung ist mit der Addition distributiv verknüpft; es gilt: $d(x \times x') = d(x) + d(x')$.

Jedes Element von $\mathfrak{R}(M \times M')$ lässt sich als $\sum (x_h \times x_h')$ darstellen; genauer gilt: Ist $(Z_1, Z_2, \cdots Z_p)$ eine additive Basis von $\mathfrak{R}(M)$—d.h. eine Basis von $\mathfrak{B}(M)$—und ist $(Z_1', Z_2', \cdots Z_{p'}')$ eine additive Basis von $\mathfrak{R}(M')$, so bilden die Elemente $Z_i \times Z_k'$ eine additive Basis in $\mathfrak{R}(M \times M')$; die Elemente von $\mathfrak{R}(M \times M')$ lassen sich in eineindeutiger Weise als $\sum t_{ik}(Z_i \times Z_k')$ mit rationalen Koeffizienten t_{ik} schreiben. (All das gilt auch für den Produktkomplex zweier Komplexe.)

Für die Multiplikation gilt [14]:

$$(x \times x') \cdot (y \times y') = (-1)^{\delta(x) \cdot \delta(y')}(x \cdot y \times x' \cdot y');$$

dabei müssen x und y' homogen-dimensional sein. Ausserdem gelten die üblichen distributiven Gesetze.

8. Der folgende Abbildungssatz für topologische Produkte wird später öfters verwendet:

A, B, A', B' seien Mannigfaltigkeiten; f sei eine Abbildung von A in A', g sei eine Abbildung von B in B'. Man konstruiert eine Abbildung h von $A \times B$ in $A' \times B'$, indem man setzt:

$$h(p \times q) = f(p) \times g(q).$$

Ist dann x, bzw. y, ein Homologieelement von A, bzw. B, so gilt:

$$h(x \times y) = f(x) \times g(y).$$

Man beweist das, indem man f und g simplizial approximiert. Ist dann a, bzw. b, ein Simplex von A, bzw. B, so ist das Bild der Zelle $a \times b$ die Zelle $f(a) \times g(b)$:

$$h(a \times b) = f(a) \times g(b);$$

denn beide Zellen rechts und links haben dieselben Eckpunkte, und auch die Orientierung stimmt überein, wie man leicht sieht. Durch Addition gewinnt man die Formel dann für Komplexe, Zyklen und Homologieklassen.

9. Wie man ganz leicht sieht, besteht die Gruppe $\mathfrak{B}(M \times M')$ der minimalen Elemente von $M \times M'$ genau aus den Elementen $(v \times z_0') + (z_0 \times v')$, wo z_0, bzw. z_0', der einfach gezählte Punkt von M, bzw. M', ist und v, bzw. v', ein beliebiges Element von $\mathfrak{B}(M)$, bzw. $\mathfrak{B}(M')$, ist.

2. Der Ring einer Gruppen-Mannigfaltigkeit

1. G sei eine geschlossene Gruppen-Mannigfaltigkeit, also eine geschlossene Mannigfaltigkeit (der Dimension n), zwischen deren Punkten eine Multiplikation erklärt ist, die die bekannten Gruppenaxiome erfüllt, und die die weitere Eigenschaft hat, dass das Produkt pq zweier Punkte p, q von G stetig von dem Paar (p, q) abhängt, und ebenso das Inverse p^{-1} stetig von p abhängt. (Gruppenmannigfaltigkeiten sind bekanntlich orientierbar.)

Ueber den Homologie-Ring von G ist der folgende Satz bekannt ([7], Nr. 2):

Satz A. *Der Ring $\mathfrak{R}(G)$ (mit rationalen Koeffizienten) einer Gruppen-Mannigfaltigkeit G ist dimensionstreu isomorph dem Ring $\mathfrak{R}(\Pi)$ eines topologischen Produktes*

$$\Pi = S_{m_1} \times S_{m_2} \times \cdots \times S_{m_l}, \qquad l \geqq 1,$$

wo S_m die m-dimensionale Sphäre bezeichnet, und die Dimensionen m_i ungerade sind.

2. Der Beweis dieses Satzes verläuft so, dass man zeigt: Bilden die Elemente $z_1, z_2, \cdots z_l$ mit der 1 zusammen ein irreduzibles Erzeugendensystem von $\mathfrak{R}(G)$, dann ist

a) das Produkt $z_1 \cdot z_2 \cdot \cdots z_l$ von Null verschieden,

b) $\delta(z_i) = m_i$ ungerade $(i = 1, \cdots, l)$ ([7], Nr. 15).

Daraus folgert man: Es ist

$$z_i \cdot z_k = -z_k \cdot z_i, \quad \text{speziell} \quad z_i \cdot z_i = 0;$$

und eine volle additive Basis von $\mathfrak{R}(G)$ (eine Basis von $\mathfrak{B}(G)$) wird gebildet von den 2^l Elementen

$$1,\ z_i\ ,\ z_{i_1}\cdot z_{i_2}\ (i_1 < i_2),\ z_{i_1}\cdot z_{i_2}\cdot z_{i_3}\ (i_1 < i_2 < i_3),\ \cdots\ ,\ z_1\cdot z_2\cdot\ \cdots\ \cdot z_l\ .$$

Diese Basis, die also durch das irreduzible Erzeugendensystem $(1, z_1, z_2, \cdots, z_l)$ bestimmt ist, heisse die Basis $[z_i]$.

Sind x, y zwei Elemente ($\neq 1$) dieser Basis, so ist $x\cdot y \neq 0$, wenn kein Element z_i in beiden Elementen zugleich als Faktor auftritt, dagegen ist $x\cdot y = 0$, wenn wenigstens ein z_i in beiden Elementen als Faktor auftritt, wie man leicht mit Hilfe der Assoziativität, der Antikommutativität der Elemente z_i und der Tatsache $z_1\cdot z_2\cdot\ \cdots\ \cdot z_l \neq 0$ bestätigt. Man kann $\mathfrak{R}(G)$ als den Ring der inhomogenen Multilinearformen in den antikommutativen Grössen $z_1, z_2, \cdots z_l$ beschreiben. Die Dimension von G ist

$$n = m_1 + m_2 + \cdots + m_l\ .$$

Das Produkt $z_1\cdot z_2\cdot\ \cdots\ \cdot z_l$ ist nulldimensional; man kann annehmen, dass es gleich dem einfach gezählten Punkt von G ist, der v_0 heissen soll (man kann nämlich etwa z_1 durch ein geeignetes rationales Vielfaches ersetzen).

Die Isomorphie zwischen $\mathfrak{R}(G)$ und $\mathfrak{R}(\Pi)$ wird folgendermassen hergestellt: Der orientierte Grundzyklus der Sphäre S_m werde auch mit S_m bezeichnet; p sei ein einfach gezählter Punkt von S_m. Man bilde die Elemente

$$\begin{aligned} Z_1 &= p \quad \times S_{m_2} \times S_{m_3} \times \cdots \times S_{m_l} \\ Z_2 &= S_{m_1} \times p \quad \times S_{m_3} \times \cdots \times S_{m_l} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ Z_l &= S_{m_1} \times S_{m_2} \times S_{m_3} \times \cdots \times p \end{aligned}$$

von $\mathfrak{R}(\Pi)$. Die Z_i haben auch die Eigenschaften a) und b); die Zuordnung $z_i \leftrightarrow Z_i$ ergibt offenbar einen Isomorphismus von $\mathfrak{R}(G)$ und $\mathfrak{R}(\Pi)$.

Die Elemente z_i sind maximal, weil sie zu einem irreduziblen Erzeugendensystem gehören (§1, Nr. 2).

Ein Produkt $z_{i_1}\cdot z_{i_2}\cdot\ \cdots\ \cdot z_{i_r}$ ist durch seine Faktoren bis auf das Vorzeichen bestimmt; in diesem Sinne kann man von *dem* Produkt der z_{i_k} sprechen, ohne auf ihre Reihenfolge zu achten.

3. Die Gruppe $\mathfrak{B}(G)$ der minimalen Elemente von $\mathfrak{R}(G)$ lässt sich leicht bestimmen. Von den Elementen der Basis $[z_i]$ sind minimal die und nur die, die Produkte von $l - 1$ Faktoren z_i sind, also die Elemente

$$v_i = z_1\cdot z_2\cdot\ \cdots\ \cdot z_{i-1}\cdot z_{i+1}\cdot\ \cdots\ \cdot z_l\ .$$

Denn ist x ein Element dieser Basis, das Produkt von höchstens $l - 2$ Faktoren z_i ist, also mindestens zwei der z_i, etwa z_r und z_s, nicht als Faktoren enthält, so ist $z_r\cdot x \neq 0$ und $0 < d(z_r\cdot x) < d(x)$, also x nicht minimal. Dagegen ist $y\cdot v_i = 0$ für jedes von z_i und 1 verschiedene Basiselement y, weil y dann mindestens einen Faktor z_k enthält, der auch in v_i vorkommt; aus $0 < \delta(y) < d(v_i)$ folgt also $y\cdot v_i = 0$ (denn wegen $\delta(z_i) = d(v_i)$ ist dann $y \neq z_i$); also ist v_i minimal. Mit derselben Ueberlegung stellt man fest, dass die von den Elementen v_i

erzeugte Untergruppe von $\mathfrak{B}(G)$ schon die ganze Gruppe $\mathfrak{B}(G)$ ist, dass also die minimalen Elemente von $\mathfrak{R}(G)$ genau die Linearkombinationen $\sum a_i v_i$ mit rationalen Koeffizienten a_i sind; die Elemente v_i $(i = 1, \cdots, l)$ bilden also eine Basis von $\mathfrak{B}(G)$.

Der Schnitt $z_i \cdot v_i$ ist gleich $\pm v_0$. Wir wollen das Vorzeichen der Elemente v_i so definieren, dass $z_i \cdot v_i = v_0$ gilt; die v_i erfüllen dann die Gleichungen

$$z_i \cdot v_k = \delta_{ik} v_0 \qquad (i, k = 1, \cdots, l),$$

durch die sie, wie man leicht überlegt, auch vollständig bestimmt sind.

Im Sphärenprodukt Π entsprechen den v_i die "Sphären selbst," nämlich die Elemente

$$\begin{aligned} V_1 &= S_{m_1} \times p \quad \times p \times \cdots \times p \\ V_2 &= p \quad \times S_{m_2} \times p \times \cdots \times p \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ V_l &= p \quad \times p \quad \times p \times \cdots \times S_{m_l}, \end{aligned}$$

die offenbar eine Basis von $\mathfrak{B}(\Pi)$ darstellen. (Um die richtigen Vorzeichen zu erhalten, muss man die Elemente Z_i aus Nr. 2 durch die Elemente $(-1)^{l-i} Z_i$ ersetzen; es gelten nämlich die Gleichungen $Z_i \cdot V_k = (-1)^{l-i} \delta_{ik} V_0$ mit $V_0 = p \times p \times \cdots \times p$. Die in Nr. 2 genannten Eigenschaften der Z_i gelten auch für die Elemente $(-1)^{l-i} Z_i$.)

Hieran sieht man unmittelbar, dass $\sum d(v_i) = n$ ist (das gilt dann natürlich für jede Basis $(v_1', v_2', \cdots v_l')$ von $\mathfrak{B}(G)$).

Die Dimension $d(v_i) = \delta(z_i) = d(S_{m_i}) = m_i$ ist ungerade; in einer Gruppe gibt es also nur ungerade-dimensionale minimale Elemente ([7], Nr. 36).

Die Zahlen l_i, die Ränge der Gruppen $\mathfrak{B}_i(G)$ (§1, Nr. 3), geben an, wieviel der Faktor-Sphären S_{m_k} des Produktes Π i-dimensional sind; durch die l_i ist Π, und folglich auch $\mathfrak{R}(G)$ vollständig bestimmt. Die Zahlen l_i sollen die charakteristischen Zahlen der Gruppe G genannt werden. l_i ist Null für gerades i. Die Zahl

$$l(G) = l_1 + l_3 + l_5 + \cdots$$

soll der Rang[4] der Gruppe heissen; der Rang einer Gruppe G ist also gleich dem Rang der Gruppe $\mathfrak{B}(G)$ der minimalen Elemente von G.

4. Die von den Elementen

$$z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r} \quad (i_1 < i_2 < \cdots < i_r), \quad 1$$

gebildete Basis $[z_i]$ ist zu sich selbst dual, d.h. die duale Basis besteht aus denselben Elementen (bis auf das Vorzeichen).

---

[4] Für den Zusammenhang dieser Definition des Ranges einer Gruppe mit der üblichen Definition des Ranges einer Lieschen Gruppe vgl.: H. Hopf: *Über den Rang geschlossener Liescher Gruppen* [Comm. Math. Helv. 13 (1940/41)].

Ist nämlich $x = z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r}$ ein beliebiges dieser Elemente, und ist $\bar{x} = z_{j_1} \cdot z_{j_2} \cdot \cdots \cdot z_{j_{l-r}}$ das Produkt derjenigen z_i, die in x nicht als Faktoren auftreten, so ist $\bar{x} \cdot x = \pm v_0$; ist aber y ein von $\bar{x}$ verschiedenes Element der Basis $[z_i]$ mit $d(y) = d(\bar{x})$, so ist $y \cdot x = 0$, weil y mindestens einen Faktor z_i enthält, der auch in x vorkommt.

In Nr. 2 sind wir ausgegangen von einem beliebigen irreduziblen Erzeugendensystem $(z_1, z_2, \cdots, z_l, 1)$ von $\mathfrak{R}(G)$; nach Nr. 3 sind den z_i die minimalen Elemente v_i zugeordnet, die eine Basis von $\mathfrak{B}(G)$ bilden. Man sieht ganz leicht, dass man auf diese Weise, d.h. durch geeignete Wahl der z_i, jede Basis von $\mathfrak{B}(G)$ erhalten kann (mit der unwesentlichen Einschränkung, dass man, wegen der Bedingung $z_1 \cdot z_2 \cdot \cdots \cdot z_l = v_0$, nur solche Basen von $\mathfrak{B}(G)$ bekommt, die durch eine lineare Transformation mit der Determinante 1 auseinander hervorgehen; das bedeutet nur, dass man eventuell v_1 durch ein rationales Vielfaches ersetzen muss.). Das beweist man z.B. mit Hilfe der Bemerkung, dass die Gleichungen $z_i \cdot v_k = \delta_{ik} v_0$, durch die bei gegebenen z_i die v_i vollständig bestimmt sind, invariant bleiben, wenn man auf die z_i und die v_i kontragrediente lineare Transformationen ausübt.

Man nehme die zu einem beliebigen irreduziblen Erzeugendensystem $(z_1, z_2, \cdots, z_l, 1)$ von $\mathfrak{R}(G)$ gehörenden minimalen Elemente v_i in eine sonst beliebige Basis von $\mathfrak{B}(G)$ auf. Dann ist das zu v_i duale Element z_i' der dualen Basis von der Form $z_i' = z_i + y_i$ mit nichtmaximalem y_i; das bestätigt man leicht auf Grund der Tatsache, dass $z_i \cdot v_i = v_0$ ist, dass aber für jedes von z_i und 1 verschiedene Element x der Basis $[z_i]$ der Schnitt $x \cdot v_i$ gleich Null ist. Diese Elemente z_i' bilden ebenfalls mit 1 zusammen ein irreduzibles Erzeugendensystem von $\mathfrak{R}(G)$; sie besitzen auch die Eigenschaften a) und b) (§1, Nr. 2); es ist nämlich $\delta(z_i') = \delta(z_i)$ und $z_1' \cdot z_2' \cdot \cdots \cdot z_l' = z_1 \cdot z_2 \cdot \cdots \cdot z_l$; aus den Gleichungen $z_i \cdot v_k = \delta_{ik} v_0$ folgen ohne weiteres die Gleichungen $z_i' \cdot v_k = \delta_{ik} v_0$.

3. Der Pontrjagin'sche Ring und der "Aufspann-Satz"

1. G sei, wie in §1, eine (geschlossene) Gruppenmannigfaltigkeit. Die Multiplikation in G, die jedem Punktepaar (p, q) von G den Punkt $p \cdot q$ zuordnet, kann man auffassen als eine Abbildung F des topologischen Produktes $G \times G$ zweier Exemplare von G in G, die definiert ist durch:

$$F(p \times q) = p \cdot q.$$

Sind x, y zwei Homologieklassen von G, so wollen wir unter dem Pontrjagin'schen Produkt [3] *(kürzer: P-Produkt) von x und y das Bild des Elementes $x \times y$ von $\mathfrak{R}(G \times G)$ bei F verstehen; wir bezeichnen es mit $x \circ y$; also:*

$$F(x \times y) = x \circ y.$$

Anschaulich, aber etwas unscharf, kann man das so sagen: Durchläuft der Punkt p den Zyklus x, und der Punkt q den Zyklus y, dann durchläuft der Punkt pq den Zyklus $x \circ y$.

Es ist $d(x \circ y) = d(x) + d(y)$.

Wegen des Assoziativgesetzes in G ist diese Multiplikation auch assoziativ; für drei Homologieelemente x_1, x_2, x_3 gilt:

$$(x_1 \circ x_2) \circ x_3 = x_1 \circ (x_2 \circ x_3).$$

Das P-Produkt ist mit der Addition distributiv verknüpft, wie man sofort sieht.

Das bedeutet: *Man kann die Betti'sche Gruppe $\mathfrak{B}(G)$ nicht nur durch die Schnittbildung $x \cdot y$ zu einem Ring $\mathfrak{R}(G)$, sondern auch durch die Pontrjagin'sche Produktbildung $x \circ y$ zu einem Ring $\mathfrak{P}(G)$, dem Pontrjagin'schen Ring von G, machen.*

Auf den Ring $\mathfrak{P}(G)$ kann man die üblichen algebraischen Begriffe anwenden; sind $x_1, x_2, \cdots x_r$ irgendwelche Elemente von $\mathfrak{P}(G)$, so ist klar, was man unter dem von den x_i erzeugten Pontrjagin'schen Teilring und was man unter dem von den x_i erzeugten Pontrjagin'schen Ideal (Rechts-, Links-, oder zweiseitig) zu verstehen hat.

(Jedes von homogen-dimensionalen x_i erzeugte Ideal in $\mathfrak{R}(G)$ ist übrigens zweiseitig, wegen der Regel $x \cdot y = \pm y \cdot x$ für homogen-dimensionale x, y; es wird sich zeigen, dass das auch in $\mathfrak{P}(G)$ so ist.)

Der Ring $\mathfrak{P}(G)$ besitzt eine Eins, nämlich das nulldimensionale Element v_0, das man etwa durch den einfach gezählten Einheitspunkt e, das Gruppen-Einselement von G, repräsentieren kann; denn wegen $F(e \times p) = ep = p$ für jeden Punkt p von G ist offenbar $F(v_0 \times x) = x$ für jedes Element x von $\mathfrak{P}(G)$, und ebenso ist $F(x \times v_0) = x$, d.h.:

$$v_0 \circ x = x \circ v_0 = x.$$

2. Der Ring $\mathfrak{P}(G)$ und sein Verhältnis zu dem Ring $\mathfrak{R}(G)$ soll nun untersucht werden.

Die Pontrjagin'schen Untersuchungen [3] haben ergeben, dass für die Gruppen der bekannten vier Killing-Cartan'schen Klassen die Ringe $\mathfrak{R}(G)$ und $\mathfrak{P}(G)$ ein ganz bestimmtes, in gewissem Sinn duales, Verhalten aufweisen. Von Cartan [15] und in anderer Form von Hopf [16] ist die Vermutung ausgesprochen worden, dass diese Dualität den Ringen $\mathfrak{R}(G)$ und $\mathfrak{P}(G)$ aller (geschlossener) Gruppenmannigfaltigkeiten zukomme. Diese Vermutung wollen wir im folgenden beweisen. Satz I ist die Cartan'sche, Satz I' die Hopf'sche Formulierung der Vermutung. Aus Gründen, die bei der Formulierung und beim Beweis der Vermutung, besonders in der Form I', klar werden werden, wollen wir die zu beweisende Tatsache den "Aufspann-Satz" nennen. Die Cartan'sche Formulierung ist [15]:

Satz I. *G sei eine geschlossene Gruppen-Mannigfaltigkeit der Dimension n, vom Rang l.*

a) *Dann gibt es im Ring $\mathfrak{P}(G)$ (mit rationalen Koeffizienten) l Elemente $v_1, v_2, \cdots v_l$ ungerader Dimensionen, für die gilt:*

$$v_i \circ v_k = -v_k \circ v_i \quad \text{und} \quad v_i \circ v_i = 0;$$

die $2^l - 1$ Elemente

$$v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} \qquad (i_1 < i_2 < \cdots < i_r)$$

bilden mit dem Punkt v_0 zusammen eine volle Betti'sche Basis (Basis der Betti'schen Gruppe $\mathfrak{B}(G)$).

b) *Sind $w_1 = v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}$ und $w_2 = v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_s}$ zwei solche Basiselemente, so ist der Schnitt $w_1 \cdot w_2$ gleich Null, wenn eines der Elemente v_i weder in dem Produkt w_1 noch in dem Produkt w_2 als Faktor vorkommt; kommt jedes v_i entweder in w_1 oder in w_2 oder in beiden als Faktor vor, dann ist der Schnitt $w_1 \cdot w_2$ (bis auf das Vorzeichen) gleich dem Produkt $v_{k_1} \circ v_{k_2} \circ \cdots \circ v_{k_t}$ $(k_1 < k_2 < \cdots < k_t)$ derjenigen v_i, die sowohl in w_1 als auch in w_2 als Faktor auftreten, oder, falls kein v_i sowohl in w_1 als auch in w_2 vorkommt, gleich v_0.*

Teil a) von Satz I beschreibt den Ring $\mathfrak{P}(G)$; die Elemente v_i sind antikommutativ und bilden ein irreduzibles Erzeugendensystem von $\mathfrak{P}(G)$; aus dem Beweis wird hervorgehen, dass man für diese Elemente v_i die Elemente einer beliebigen Basis der Gruppe $\mathfrak{B}(G)$ der minimalen Elemente von $\mathfrak{R}(G)$ wählen darf. Die in I a) genannte Basis von $\mathfrak{B}(G)$ wollen wir die Basis $\langle v_i \rangle$ nennen; sie ist durch die Elemente v_i bestimmt.

Teil b) enthält die Beziehung zwischen $\mathfrak{R}(G)$ und $\mathfrak{P}(G)$; denn b) bestimmt den Schnitt zweier beliebiger Elemente von $\mathfrak{P}(G)$. Diese Beziehung tritt klarer hervor in der Hopf'schen Formulierung des Satzes I, die durch den Satz I' gegeben wird. Dazu ist eine Vorbereitung nötig.

3. $\Pi = S_{m_1} \times S_{m_2} \times S_{m_3} \times \cdots \times S_{m_l}$ sei ein Produkt von Sphären S_{m_i} der Dimensionen m_i. Z_i und V_i seien die in §2, Nr. 2 u. 3 definierten maximalen und minimalen Elemente von $\mathfrak{R}(\Pi)$.

In $\mathfrak{B}(\Pi)$ lässt sich neben der Schnittbildung in naheliegender Weise noch eine zweite Multiplikation, das "Aufspannen," erklären [16]. Das Produkt zweier Elemente X, Y von $\mathfrak{B}(\Pi)$ wird mit $X \boxtimes Y$ bezeichnet; die Definition geschieht folgendermassen:

Für $i_1 < i_2 < \cdots < i_r$ ist $V_{i_1} \boxtimes V_{i_2} \boxtimes \cdots \boxtimes V_{i_r}$ dasjenige Element $x_1 \times x_2 \times \cdots \times x_l$ von $\mathfrak{B}(\Pi)$, in dem

$$x_j = S_{m_j} \quad \text{für} \quad j = i_1, i_2, \cdots, i_r \text{ und}$$
$$x_j = p \qquad \text{für alle anderen } j$$

ist. Das sind mit dem Punkt V_0 zusammen 2^l Elemente, die eine volle Betti'sche Basis von $\mathfrak{B}(\Pi)$ bilden.

Für zwei solche Elemente

$$X = V_{i_1} \boxtimes V_{i_2} \boxtimes \cdots \boxtimes V_{i_r}, \qquad Y = V_{j_1} \boxtimes V_{j_2} \boxtimes \cdots \boxtimes V_{j_s}$$

wird das "aufgespannte" Element $X \boxtimes Y$ dadurch definiert, dass man die rechten Seiten nach dem assoziativen Gesetz miteinander multipliziert und die Regeln

$$V_i \boxtimes V_k = (-1)^{m_i \cdot m_k} V_k \boxtimes V_i, \qquad V_i \boxtimes V_i = 0$$

$(m_i = d(V_i))$ anwendet; dann wird $X \boxtimes Y$ entweder $= 0$ oder (bis auf das Vorzeichen) eins der oben genannten Basiselemente. Für beliebige Elemente

von $\mathfrak{B}(\Pi)$ definiert man das Produkt, indem man sie durch die Elemente dieser Basis darstellt und die distributiven Gesetze anwendet. Es ist $d(X \boxtimes Y) = d(X) + d(Y)$.

Durch diese Multiplikation wird die Betti'sche Gruppe $\mathfrak{B}(\Pi)$ zu einem Ring $\mathfrak{Q}(\Pi)$. Es existiert eine Eins:

$$V_0 = p \times p \times \cdots \times p.$$

Die Ringe $\mathfrak{R}(\Pi)$ und $\mathfrak{Q}(\Pi)$ sind isomorph; die Isomorphie bekommt man, wenn man die Erzeugenden Z_i von $\mathfrak{R}(\Pi)$ den Erzeugenden V_i von $\mathfrak{Q}(\Pi)$ zuordnet.

Es besteht aber weiter eine gewisse Dualität; es gilt nämlich die folgende "Formel (Δ)":

$$(\Delta) \qquad Z_{i_1} \cdot Z_{i_2} \cdot \cdots \cdot Z_{i_r} = \pm V_{j_1} \boxtimes V_{j_2} \boxtimes \cdots \boxtimes V_{j_{l-r}},$$

wenn die Mengen $\{i_1, i_2, \cdots, i_r\}$ und $\{j_1, j_2, \cdots, j_{l-r}\}$ komplementäre Teilmengen der Menge der Zahlen von 1 bis l sind. Beide Seiten bedeuten nämlich dasjenige Element $x_1 \times x_2 \times \cdots \times x_l$ von $\mathfrak{B}(\Pi)$, in dem

$$x_j = S_{m_j} \quad \text{für} \quad j = j_1, j_2, \cdots, j_{l-r} \text{ und}$$
$$x_j = p \quad \text{für} \quad j = i_1, i_2, \cdots, i_r$$

ist. Das Vorzeichen hängt auf gewisse Weise von den Indizes ab.

Mit Hilfe dieser Formel (Δ) lässt sich jedes Element von $\mathfrak{B}(\Pi)$ sowohl als Element von $\mathfrak{R}(\Pi)$ als auch als Element von $\mathfrak{Q}(\Pi)$ schreiben. Infolgedessen kann man den Schnitt $X \cdot Y$ zweier beliebiger Elemente X, Y von $\mathfrak{Q}(\Pi)$ und das Produkt $X \boxtimes Y$ zweier beliebiger Elemente X, Y von $\mathfrak{R}(\Pi)$ berechnen; die Formel (Δ) bestimmt die Beziehung zwischen $\mathfrak{R}(\Pi)$ und $\mathfrak{Q}(\Pi)$ vollkommen.

4. Der Satz I′ lautet nun folgendermassen [16]:

Satz I′. *Die auf Grund von Satz A mögliche, isomorphe Abbildung der Ringe* $\mathfrak{R}(G)$ *und* $\mathfrak{R}(\Pi)$ (*wo* Π *ein gewisses Sphärenprodukt ist*) *lässt sich so wählen, dass zugleich die Ringe* $\mathfrak{P}(G)$ *und* $\mathfrak{Q}(\Pi)$ *isomorph aufeinander abgebildet werden.*

Der Beweis für die Sätze I und I′ wird in den Nummern 5–8 geliefert; er geht folgendermassen vor sich: Wir beweisen zunächst (in Nr. 5, 6, 7) den Teil a) von Satz I, und zwar in der folgenden verschärften Form: Die Elemente v_i einer beliebigen Basis $(v_1, v_2, \cdots v_l)$ der Gruppe $\mathfrak{B}(G)$ der minimalen Elemente von $\mathfrak{R}(G)$ haben die in I a) genannten Eigenschaften. Den Teil b) von Satz I beweisen wir in Nr. 8 folgendermassen: Wir bilden mit den Elementen v_i die in I a) genannte Basis $\langle v_\iota \rangle$; die in der dualen Basis zu den v_i dualen Elemente nennen wir z_i; sie haben die in §2, Nr. 4, Ende, genannten Eigenschaften. Dann beweisen wir die Gültigkeit der der Formel (Δ) (Nr. 3) entsprechenden "Formel (D)":

$$(\mathrm{D}) \qquad z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r} = \pm v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_{l-r}},$$

für komplementäre Teilmengen $\{i_1, i_2, \cdots, i_r\}$ und $\{j_1, j_2, \cdots, j_{l-r}\}$ der Menge $\{1, 2, \cdots, l\}$.

Aus dieser Formel folgt dann sofort I b): Man schreibe die Elemente w_1 und w_2 mit Hilfe von (D) als Schnitt gewisser z_i; dann bilde man den Schnitt $w_1 \cdot w_2$, der entweder 0 oder wieder ein Schnitt gewisser z_i ist; im zweiten Falle schreibt man $w_1 \cdot w_2$ wieder mit Hilfe von (D) als P-Produkt gewisser v_i. Man bestätigt ohne Mühe, dass der Schnitt $w_1 \cdot w_2$ das in I b) angegebene Verhalten zeigt, wegen der Dualität der Formel (D), die man folgendermassen aussprechen kann: Kommt in (D) auf der linken Seite das Element z_i vor, so kommt rechts das dazu duale v_i nicht vor; kommt links das Element z_k nicht vor, so kommt rechts das dazu duale v_k vor.

Der Satz I′ ist dann auch bewiesen: ordnet man die eben genannten z_i den Elementen Z_i von $\mathfrak{R}(\Pi)$ zu, so erhält man einen Isomorphismus $\mathfrak{R}(G) \approx \mathfrak{R}(\Pi)$. Dabei entsprechen wegen §2, Nr. 4 die Elemente v_i von $\mathfrak{R}(G)$ den Elementen V_i von $\mathfrak{R}(\Pi)$. Wegen der Formeln (D) und (Δ) entspricht dann dem P-Produkt irgendwelcher v_{i_k} das von den entsprechenden V_{i_k} "aufgespannte" Element, d. h. es liegt ein Isomorphismus zwischen $\mathfrak{P}(G)$ und $\mathfrak{Q}(\Pi)$ vor; um Uebereinstimmung der Vorzeichen in (D) und (Δ) zu erhalten, muss man jedoch die Elemente Z_i durch die Elemente $(-1)^{l-i}Z_i$ ersetzen (vgl. §2, Nr. 3).

5. Wir beginnen mit dem Beweis von I und I′. Wir betrachten die Abbildung

$$F(p \times q) = pq$$

des Produktes $G \times G$ zweier Exemplare von G in G; ϕ sei der zugehörige Umkehrungshomomorphismus (§1, Nr. 5).

Das Bild $\phi(z)$ eines beliebigen Elementes z von $\mathfrak{R}(G)$ hat eine sehr spezielle Gestalt, auf der alle weiteren Ueberlegungen beruhen; sie ist eine Folge der Tatsache, dass der Punkt v_0 in $\mathfrak{P}(G)$ die Rolle der Eins spielt, also eine Folge der Gleichungen

$$F(x \times v_0) = x, \qquad F(v_0 \times y) = y$$

für beliebige Elemente x, y von $\mathfrak{B}(G)$.

Das Bild $\phi(z)$ eines beliebigen $z(\neq 1)$ lässt sich in der folgenden Form schreiben:

$$\phi(z) = (1 \times y) + (x \times 1) + \sum (x_h \times y_h),$$

wo die Ungleichungen $d(z) < d(x_h) < n$, $d(z) < d(y_h) < n$ für alle h gelten; denn auf die Form der rechten Seite kann man jedes Element von $\mathfrak{R}(G \times G)$ bringen, und die Ungleichungen folgen aus der Gleichung $\delta(\phi(z)) = \delta(z)$. Hierin lassen sich nun y und x bestimmen:

Wir bilden den Schnitt $\phi(z) \cdot (v_0 \times 1)$. Der Schnitt von v_0 mit nicht skalaren Elementen ist immer 0; auf Grund der in §1, Nr. 7 genannten Sätze findet man sofort:

$$\phi(z) \cdot (v_0 \times 1) = v_0 \times y.$$

Uebt man auf diese Gleichung die Abbildung F aus und beachtet die Funktionalgleichung für ϕ (§1, Nr. 5), so erhält man:

$$z \cdot 1 = y, \quad \text{also} \quad z = y.$$

Entsprechend bilden wir den Schnitt $\phi(z) \cdot (1 \times v_0)$. Man findet:

$$\phi(z) \cdot (1 \times v_0) = (-1)^{\delta(z) \cdot \delta(v_0)} (x \times v_0).$$

Durch Ausübung von F und Anwendung der Funktionalgleichung für ϕ erhält man, wenn man noch beachtet, dass $\delta(v_0) = n$ ist:

$$z = (-1)^{n \cdot \delta(z)} x.$$

Damit ist gezeigt: Für jedes Element $z (\neq 1)$ von $\mathfrak{R}(G)$ gilt die "Formel (ϕ)":

$$(\phi) \qquad \phi(z) = (1 \times z) + (-1)^{n \cdot \delta(z)} (z \times 1) + \sum (x_h \times y_h)$$

mit den Ungleichungen $d(z) < d(x_h) < n$, $d(z) < d(y_h) < n$. Ausserdem ist natürlich $\phi(1) = 1 \times 1$.

Wir führen nun eine Basis in $\mathfrak{B}(G)$ ein. Und zwar wählen wir die von einem beliebigen irreduziblen Erzeugendensystem $(1, z_1, z_2, \cdots z_l)$ von $\mathfrak{R}(G)$ erzeugte Basis $[z_i]$ (§2, Nr. 2); wir dürfen $z_1 \cdot z_2 \cdot \cdots \cdot z_l = v_0$ annehmen. Das Vorzeichen der Elemente v_i, die zu dieser Basis gehören, ist durch

$$v_i = (-1)^{i-1} z_1 \cdot z_2 \cdot \cdots \cdot z_{i-1} \cdot z_{i+1} \cdot \cdots \cdot z_l$$

definiert; dann sind in der dualen Basis, die nach §2, Nr. 4 bis auf Vorzeichen aus denselben Elementen wie die Basis $[z_i]$ besteht, gerade die Elemente z_i zu den Elementen v_i dual, d. h. es ist $z_i \cdot v_i = v_0$; es gelten dann die Gleichungen:

$$z_i \cdot v_k = \delta_{ik} v_0 .$$

Die Elemente v_i bilden eine Basis von $\mathfrak{B}(G)$; nach §2, Nr. 4 kann man durch geeignete Wahl der Elemente z_i jede Basis von $\mathfrak{B}(G)$ auf diese Weise erhalten.

Durch die Basis $[z_i]$ in $\mathfrak{B}(G)$ ist eine Basis in $\mathfrak{B}(G \times G)$ bestimmt (§1, Nr. 7), die wir die Basis $[z_i]^{\times}$ nennen wollen; auch diese Basis ist (bis auf Vorzeichen) zu sich dual.

Für die maximalen Elemente z_i ist $\delta(z_i)$ ungerade. Daher heisst für sie die Formel (ϕ):

$$(\phi_i) \qquad \phi(z_i) = (1 \times z_i) + (-1)^n (z_i \times 1) + R_i ;$$

dabei ist der Term $\sum (x_h \times y_h)$ aus (ϕ) mit R_i bezeichnet worden. Weil alle $\delta(z_i)$ ungerade sind, ist für jeden Term $(x_h \times y_h)$ von R_i entweder $\delta(x_h)$ oder $\delta(y_h)$ gerade, also nie sowohl x_h als auch y_h maximal, sondern wenigstens eines zusammengesetzt.

6. Wir beweisen zunächst die Behauptungen

$$v_i \circ v_i = 0 \quad \text{und} \quad v_i \circ v_k = -v_k \circ v_i$$

von I a); anders gesagt, wir beweisen: für die Elemente $v_i \times v_i$ und $(v_i \times v_k) + (v_k \times v_i)$, die wir für den Moment mit v_{ii} und v_{ik} bezeichnen, gilt:

$$F(v_{ii}) = F(v_{ik}) = 0.$$

Zum Beweis beachten wir, dass die Funktionalgleichung für ϕ die Gleichung $F(\phi(x')\cdot v_{ii}) = x'\cdot F(v_{ii})$ und die entsprechende für v_{ik} mit einem beliebigen $x' \in \mathfrak{B}(G)$ liefert. Um zu zeigen, dass $F(v_{ii}) = 0$ ist, genügt es wegen des Poincaré-Veblen'schen Dualitätssatzes, die folgende Behauptung (a) zu beweisen:

(a) für jedes x von $\mathfrak{B}(G)$ mit $\delta(x) = d(v_{ii})$ ist die Schnittzahl $\phi(x)\cdot v_{ii}$ gleich Null.

Ebenso genügt es für den Beweis von $F(v_{ik}) = 0$, zu zeigen:

(b) für jedes y von $\mathfrak{B}(G)$ mit $\delta(y) = d(v_{ik})$ ist die Schnittzahl $\phi(y)\cdot v_{ik}$ gleich Null.

Wir betrachten nun die Menge derjenigen Elemente $\sum (x_h \times y_h)$ in $\mathfrak{R}(G \times G)$, in denen in jedem Term $x_h \times y_h$ entweder x_h oder y_h oder beide zusammengesetzte Elemente von $\mathfrak{R}(G)$ sind; wie man sofort einsieht, ist diese Menge, die $\mathfrak{U}$ heisse, ein zweiseitiges Ideal in $\mathfrak{R}(G \times G)$. Die Elemente $(z_i \times z_i)$, $(z_i \times z_k)$, $(z_k \times z_i)$ gehören zu verschiedenen Restklassen von $\mathfrak{U}$.

Aus der Formel (ϕ_i) und der anschliessenden Bemerkung über R_i folgen die Kongruenzen:

$$\phi(z_r) \equiv (1 \times z_r) + (-1)^n(z_r \times 1) \qquad \text{mod } \mathfrak{U}.$$

Weil die Elemente $(1 \times z_r\cdot z_s)$ und $(z_r\cdot z_s \times 1)$ zu $\mathfrak{U}$ gehören, folgt daraus sofort:

$$\phi(z_r\cdot z_s) = \phi(z_r)\cdot\phi(z_s) \equiv (-1)^n(z_s \times z_r) - (-1)^n(z_r \times z_s) \qquad \text{mod } \mathfrak{U},$$

wo das Vorzeichen des zweiten Gliedes daraus folgt, dass die $\delta(z_j)$ ungerade sind. Ebenso rechnet man sofort aus:

$$\phi(z_r\cdot z_s\cdot z_t) = \phi(z_r)\cdot\phi(z_s)\cdot\phi(z_t) \equiv 0 \qquad \text{mod } \mathfrak{U},$$

und das Gleiche für alle Produkte mit mehr als drei Faktoren. Stellt man also für irgendein Element x der Basis $[z_i]$ von $\mathfrak{B}(G)$ das Bild $\phi(x)$ durch die Basis $[z_i]^{\times}$ von $\mathfrak{B}(G \times G)$ dar, so tritt darin das Basiselement $z_i \times z_i$ nicht (d. h. mit dem Koeffizienten Null) auf; die Basiselemente $z_i \times z_k$ und $z_k \times z_i$ treten nur dann mit nichtverschwindendem Koeffizienten auf, wenn $x = z_i\cdot z_k$ ist, und dann treten sie mit entgegensetzten Koeffizienten auf (nämlich mit $+1$ und -1).

Nun sind aber die Elemente $(z_i \times z_i)$, $(z_i \times z_k)$, $(z_k \times z_i)$, wenn man sie noch mit den Vorzeichen $(-1)^n$ versieht, die zu den Elementen $(v_i \times v_i)$, $(v_i \times v_k)$, $(v_k \times v_i)$ dualen Elemente der Basis $[z_i]^{\times}$ (man beachte, dass Basis und duale Basis bis auf Vorzeichen mit der Basis $[z_i]^{\times}$ übereinstimmen).

Daraus schliesst man sofort die Behauptung (a):

$$\phi(x)\cdot(v_i \times v_i) = 0, \quad \text{wenn} \quad \delta(x) = d(v_i \times v_i).$$

Auch für die Behauptung (b) muss man nur noch den Schnitt $\phi(z_i\cdot z_k)\cdot v_{ik}$ prüfen. Aber wegen der Verschiedenheit der Vorzeichen von $(z_i \times z_k)$ und $(z_k \times z_i)$ in $\phi(z_i\cdot z_k)$ ist auch dieser Schnitt Null. Damit ist (a) und (b), also auch die Behauptung am Anfang dieser Nummer, bewiesen.

7. Als zweites beweisen wir die Gleichung:

$$v_1 \circ v_2 \circ \cdots \circ v_l = 1.$$

Wegen der eben bewiesenen Antikommutativität der v_i dürfen wir annehmen, die v_i seien so angeordnet, dass $d(v_i) \leqq d(v_k)$ ist für $i < k$.

Dass das Produkt der v_i gerade $+1$ ist, liegt daran, dass die z_i so definiert sind, dass $z_1 \cdot z_2 \cdots z_l = +v_0$ ist; das Wesentliche ist, dass das—n-dimensionale—Element $v_1 \circ v_2 \circ \cdots \circ v_l$ nicht Null ist, d.h. dass die ganze Mannigfaltigkeit G von der Basis $\{v_1, v_2, \cdots v_l\}$ der Gruppe $\mathfrak{B}(G)$ "aufgespannt" wird; dem entspricht die Gleichung $V_1 \boxtimes V_2 \boxtimes \cdots \boxtimes V_l = 1$ beim Sphärenprodukt Π.

Wir bezeichnen mit w_k das Teilprodukt $v_k \circ v_{k+1} \circ \cdots \circ v_l$ ($k = 1, 2, \cdots, l$) und setzen noch $w_{l+1} = v_0$.

Dann beweisen wir für $k = 1, 2, \cdots, l$ die Gleichungen:

$$(1_k) \qquad z_k \cdot w_k = (-1)^{l-k} w_{k+1}$$

$$(2_k) \qquad z_i \cdot w_k = 0 \qquad \text{für } i < k\cdot$$

Das ist richtig für $k = l$, wegen $z_l \cdot v_l = v_0$. Die Gleichungen seien bewiesen für k; wir beweisen sie für $k - 1$.

Wir gehen aus von der Tatsache

$$F(v_{k-1} \times w_k) = w_{k-1},$$

die aus der Definition der w_k folgt.

Für ein beliebiges Element z_i ergibt die Funktionalgleichung für ϕ die Gleichung:

$$(3) \qquad F(\phi(z_i) \cdot (v_{k-1} \times w_k)) = z_i \cdot w_{k-1}.$$

Für $i < k$ setze man jetzt $\phi(z_i)$ aus der Formel (ϕ_i) (Nr. 5) ein. Man erhält:

$$(4) \qquad \phi(z_i) \cdot (v_{k-1} \times w_k) = (-1)^{n+k-1}((z_i \cdot v_{k-1}) \times w_k) \qquad \text{für } i < k.$$

Denn wegen (2_k) ist $(1 \times z_i) \cdot (v_{k-1} \times w_k) = 0$; und ist $x \times y$ ein Element aus $R_i = \sum x_h \times y_h$, so ist wegen $d(x) > d(z_i) \geqq d(z_{k-1})$ der Schnitt $x \cdot v_{k-1} = 0$, also auch $R_i \cdot (v_{k-1} \times w_k) = 0$ (die zweite Ungleichung folgt aus der Voraussetzung über die $d(v_i)$).

Für $i = k - 1$ erhält man also:

$$(1') \qquad \phi(z_{k-1}) \cdot (v_{k-1} \times w_k) = (-1)^{l-(k-1)}(v_0 \times w_k),$$

für $i < k - 1$ erhält man dagegen:

$$(2') \qquad \phi(z_i) \cdot (v_{k-1} \times w_k) = 0 \qquad \text{für } i < k - 1.$$

Aus $(1')$ und (3) entnimmt man jetzt:

$$z_{k-1} \cdot w_k = (-1)^{l-(k-1)} w_k,$$

aus $(2')$ und (3) dagegen:

$$z_i \cdot w_{k-1} = 0 \qquad \text{für } i < k - 1;$$

und das sind die Gleichungen (1_{k-1}) und (2_{k-1}).

Aus den damit für $k = 1, 2, \cdots, l$ bewiesenen Gleichungen (1_k) schliesst man jetzt der Reihe nach: $w_l \neq 0$ (wegen $w_{l+1} = v_l \neq 0$), $w_{l-1} \neq 0, \cdots, w_1 \neq 0$. Also ist jedenfalls $w_1 = v_1 \circ v_2 \circ \cdots \circ v_l \neq 0$. Bei Durchführung der elementaren Rechnung erhält man die genauere Aussage:

$$v_1 \circ v_2 \circ \cdots \circ v_l = 1.$$

Aus den Tatsachen, dass die v_i antikommutativ sind (in $\mathfrak{P}(G)$), und dass $v_1 \circ v_2 \circ \cdots \circ v_l \neq 0$ ist, kann man nun schliessen, dass die Elemente v_i (zusammen mit v_0) den Ring $\mathfrak{P}(G)$ erzeugen. Dazu betrachten wir die 2^l Elemente

$$v_0 \quad \text{und} \quad v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} \qquad (i_1 < i_2 < \cdots < i_r).$$

Wegen der Antikommutativität der v_i ist jedes beliebige Potenz-P-Produkt, das man aus den v_i bilden kann, entweder 0 oder—bis aufs Vorzeichen—einem dieser Elemente gleich. Nun stellt man leicht fest, dass diese Elemente linear unabhängig sind:

$0 = \sum a_i \cdot y_i$ sei eine Relation zwischen solchen Elementen y_i mit Koeffizienten a_i. Es sei y_1 ein "kürzestes" dieser Elemente, d. h. eines mit möglichst kleiner Zahl von Faktoren v_i; mit $\bar{y}_1$ werde das P-Produkt derjenigen v_i bezeichnet, die in y_1 nicht als Faktoren auftreten.

Nun multipliziere man die Gleichung $0 = \sum a_i \cdot y_i$ im Sinne des P-Produktes mit $\bar{y}_1$. Auf der rechten Seite ergibt $a_1 \cdot y_1 \circ \bar{y}_1$ das Element $\pm a_1 \cdot 1$; denn es ist $y_1 \circ \bar{y}_1 = \pm v_1 \circ v_2 \circ \cdots \circ v_l$. Für die y_i mit $i > 1$ ist aber $y_i \circ \bar{y}_1 = 0$, weil mindestens ein v_k in y_i und in $\bar{y}_1$ zugleich vorkommt. Man erhält also: $0 = \pm a_1$; ebenso zeigt man, dass alle a_i verschwinden.

Nun hat aber $\mathfrak{B}(G)$ den Rang 2^l (§2, Nr. 2). Also bilden die Elemente

$$v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} \qquad (i_1 < i_2 < \cdots < i_r)$$

mit v_0 zusammen eine Basis von $\mathfrak{B}(G)$, die die Basis $\langle v_i \rangle$ genannt werden soll.

Da nach §2, Nr. 3 die $d(v_i)$ ungerade sind, ist jetzt I a) vollständig bewiesen.

Die Elemente v_i bilden ein irreduzibles Erzeugendensystem von $\mathfrak{P}(G)$.

8. Wir kommen jetzt zum Beweis von I b). Wir betrachten wieder die Abbildungen F und ϕ. Aber als Basis in G und zur Bildung der Basis in $G \times G$ nehmen wir jetzt die Basis $\langle v_i \rangle$. In der dualen Basis sind zu den v_i dual gewisse Elemente z_i', die nach §2, Nr. 4 ein irreduzibles Erzeugendensystem von $\mathfrak{R}(G)$ bilden, und für die die Gleichungen

$$z_i' \cdot v_k = \delta_{ik} v_0$$

gelten. Die z_i' nennen wir wieder z_i; für die so durch die v_i bestimmten z_i beweisen wir die Formel (D).

Wir behaupten: für die so bestimmten z_i verschwindet der Rest R_i in der Formel (ϕ_i), d. h. es gilt die verfeinerte Formel (ϕ_0):

$$(\phi_0) \qquad \phi(z_i) = (1 \times z_i) + (-1)^n (z_i \times 1).$$

Zum Beweis betrachten wir die Abbildung F; sie ist vollständig zu übersehen. Die Bilder der Elemente der durch die Basis $\langle v_i \rangle$ von $\mathfrak{B}(G)$ bestimmten Basis $\langle v_i \rangle^\times$ von $\mathfrak{B}(G \times G)$ werden durch folgende Gleichungen gegeben:

$$F((v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}) \times (v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_s})) = v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} \circ v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_s} .$$

Dabei ist das Element rechts entweder 0 oder—bis aufs Vorzeichen—ein Element der Basis $\langle v_i \rangle$. ($v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}$ oder $v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_s}$ kann auch das Element v_0 bedeuten; dann ist F durch $F(v_0 \times x) = F(x \times v_0) = x$ gegeben.)

Nun seien w_1 und w_2 zwei beliebige Elemente der Basis $\langle v_i \rangle$ von $\mathfrak{B}(G)$, die der Bedingung $\delta(z_i) = d(w_1) + d(w_2)$ genügen, sodass also $w = w_1 \times w_2$ ein Element der Basis $\langle v_i \rangle^\times$ von $\mathfrak{B}(G \times G)$ mit $\delta(\phi(z_i)) = d(w)$ ist. Wir betrachten die Schnittzahl $\phi(z_i) \cdot w$. Wegen der Funktionalgleichung für ϕ ist

$$F(\phi(z_i) \cdot w) = z_i \cdot F(w).$$

Weil z_i das zu v_i duale Element ist, ist diese Schnittzahl nur dann nicht Null, wenn $F(w)$, das ja ein Element der Basis $\langle v_i \rangle$ von $\mathfrak{B}(G)$ ist, gleich v_i ist. Daraus folgt: Der Schnitt $\phi(z_i) \cdot w$ ist nur dann nicht Null, wenn entweder $w = v_i \times v_0$ oder $w = v_0 \times v_i$ ist. Das bedeutet: in der Darstellung von $\phi(z_i)$ durch die zu der Basis $\langle v_i \rangle^\times$ duale Basis treten nur die zu $v_i \times v_0$ und $v_0 \times v_i$ dualen Elemente mit von Null verschiedenem Koeffizienten auf; und das sind eben die Elemente $(1 \times z_i)$ und $(-1)^n(z_i \times 1)$. Damit ist die Formel (ϕ_0) bewiesen.

Wir betrachten ein beliebiges Produkt $v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}$ (mit lauter verschiedenen Faktoren v). Wir behaupten das Bestehen der folgenden Gleichungen:

$$(1) \quad z_{i_1} \cdot (v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}) = \begin{cases} v_0 & \text{für } r = 1, \\ (-1)^{r-1} v_{i_2} \circ v_{i_3} \circ \cdots \circ v_{i_r} & \text{für } r > 1, \end{cases}$$

$$(2) \quad z_k \cdot (v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}) = 0, \text{ wenn keiner der Indizes } i_1, i_2, \cdots, i_r \text{ gleich } k \text{ ist}$$

Beachtet man die Antikommutativität der v_i, so erhält man aus der Gleichung (1) die folgende etwas allgemeinere, in der i_k ein beliebiger der Indizes i_1, $i_2, \cdots, i_r$ ist:

$$(1') \quad z_{i_k} \cdot (v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_k} \circ \cdots \circ v_{i_r})$$
$$= \pm v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_{k-1}} \circ v_{i_{k+1}} \circ \cdots \circ v_{i_r} \quad (\text{für } r > 1).$$

Diese Gleichungen besagen: Der Schnitt eines Elementes $w = v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}$ mit einem Element z_j ist dann und nur dann nicht Null, wenn das zu z_j duale v_j als Faktor in w auftritt; tritt v_j als Faktor in w auf, dann erhält man den Schnitt $z_j \cdot w$ (bis aufs Vorzeichen), indem man den Faktor v_j aus dem Produkt $v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}$ weglässt.

Den Beweis führen wir durch Induktion nach r; für $r = 1$ ist die Richtigkeit von (1) und (2) klar. (1) und (2) seien also richtig für alle P-Produkte mit weniger als r Faktoren v_i.

Wir gehen aus von der Tatsache:

$$F(v_{i_1} \times (v_{i_2} \circ v_{i_3} \circ \cdots \circ v_{i_r})) = v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r},$$

die aus der Definition des P-Produktes folgt.

Nach der Funktionalgleichung für ϕ ist:

$$(3) \qquad F(\phi(z_j) \cdot (v_{i_1} \times (v_{i_2} \circ v_{i_3} \circ \cdots \circ v_{i_r}))) = z_j \cdot (v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}).$$

In dem darin auftretenden Produkt

$$\phi(z_j) \cdot (v_{i_1} \times (v_{i_2} \circ v_{i_3} \circ \cdots \circ v_{i_r})),$$

das wir für den Moment X nennen wollen, setzen wir für $\phi(z_j)$ den Ausdruck aus Formel (ϕ_0) ein. Wir erhalten:

$$X = (v_{i_1} \times (z_j \cdot (v_{i_2} \circ v_{i_3} \circ \cdots \circ v_{i_r}))) + (-1)^{r-1}((z_j \cdot v_{i_1}) \times (v_{i_2} \circ v_{i_3} \circ \cdots \circ v_{i_r})).$$

Ist erstens $j = i_1$, so ist nach Induktionsvoraussetzung $z_j \cdot (v_{i_2} \circ v_{i_3} \circ \cdots \circ v_{i_r})$ gleich Null; andererseits ist dann $z_j \cdot v_{i_1} = v_0$. Also ist

$$X = (-1)^{r-1} v_0 \times (v_{i_2} \circ v_{i_3} \circ \cdots \circ v_{i_r})$$

und die Gleichung (3) liefert die Gleichung (1) für den Index r.

Ist zweitens j keiner der Indizes $i_1, i_2, \cdots, i_r$, so ist nach Induktionsvoraussetzung $z_j \cdot (v_{i_2} \circ v_{i_3} \circ \cdots \circ v_{i_r})$ gleich Null. Aber auch $z_j \cdot v_{i_1}$ ist gleich Null; also ist $X = 0$. Dann liefert die Gleichung (3) die Gleichung (2) für den Index r.

Durch die Gleichungen (1) und (2) ist das gegenseitige Verhalten von $\mathfrak{R}(G)$ und $\mathfrak{P}(G)$ vollkommen bestimmt; wir wollen jetzt die Dualitätsformel (D) daraus ableiten:

$$(\text{D}) \qquad z_{i_1} \cdot z_{i_2} \cdots z_{i_r} = \pm v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_{l-r}},$$

wenn $\{i_1, i_2, \cdots, i_r\}$ und $\{j_1, j_2, \cdots, j_{l-r}\}$ komplementäre Teilmengen der Menge der Zahlen von 1 bis l sind.

(Das Vorzeichen hängt auf etwas komplizierte Weise von den Indizes ab; es sei bemerkt, dass

$$z_1 \cdot z_2 \cdots z_r = (-1)^{r(l-r)} v_{r+1} \circ v_{r+2} \circ \cdots \circ v_l$$

ist. Dass die Vorzeichen in (D) und (Δ) übereinstimmen (nach der Ersetzung von Z_i durch $(-1)^{l-i} Z_i$, vgl. Nr. 4), ergibt sich nach einer elementaren Rechnung.)

Wir gehen aus von der Tatsache

$$v_1 \circ v_2 \circ \cdots \circ v_l = 1,$$

und betrachten ein beliebiges Produkt $z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r}$. Wir multiplizieren die Gleichung $v_1 \circ v_2 \circ \cdots \circ v_l = 1$ mit z_{i_1}. Rechts entsteht dabei z_{i_1}; und links fällt nach Gleichung (1') der Faktor v_{i_1} weg. Die so entstandene Gleichung multiplizieren wir mit z_{i_2}. Rechts entsteht dabei $z_{i_1} \cdot z_{i_2}$; und links fällt auch noch

der Faktor v_{i_2} weg. So fahren wir fort; jedesmal kommt rechts ein Faktor z_{i_k} (im Schnitt-Sinn) dazu, und links fällt der P-Faktor v_{i_k} weg. So entsteht die Gleichung (D). Nach Nr. 5 ist damit 1 b), und damit der ganze "Aufspann-Satz" bewiesen.

9. Zum Rechnen in $\mathfrak{P}(G)$ bemerken wir noch folgendes:
Gehen die Elemente $v'_{i_1}, v'_{i_2}, \cdots, v'_{i_r}$ aus den Elementen $v_{i_1}, v_{i_2}, \cdots, v_{i_r}$ durch eine lineare Transformation

$$v'_{i_j} = \sum a_{jk} v_{i_k} \qquad (j, k = 1, 2, \cdots, r)$$

mit der Determinante $|a_{jk}|$ hervor, so gilt wegen der Antikommutativität der v_i die Gleichung:

$$v'_{i_1} \circ v'_{i_2} \circ \cdots \circ v'_{i_r} = |a_{jk}| v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}.$$

Daraus folgt z. B., dass das Produkt (in $\mathfrak{P}(G)$) von linear abhängigen minimalen Elementen $v'_{i_1}, v'_{i_2}, \cdots, v'_{i_r}$ immer Null ist.

Weiter gilt: für beliebige (homogen-dimensionale) Elemente x, y ist

$$x \circ y = (-1)^{d(x) \cdot d(y)} y \circ x;$$

zum Beweis muss man nur x und y in der Basis $\langle v_i \rangle$ schreiben.

In Nr. 5 sind wir ausgegangen von einer beliebigen Basis $[z_i]$ von $\mathfrak{B}(G)$; dadurch wurden die Elemente v_i, die eine Basis von $\mathfrak{B}(G)$ bilden, bestimmt; nach §2, Nr. 4 kann man jede Basis von $\mathfrak{B}(G)$ auf diese Weise bekommen. In der Fassung I des Aufspannsatzes kann man also unter den Elementen v_i die Elemente einer beliebigen Basis von $\mathfrak{B}(G)$ verstehen. Die Elemente z_i, die dann in der Formel (D) auftreten (und die ein irreduzibles Erzeugendensystem von $\mathfrak{R}(G)$ bilden), sind folgendermassen bestimmt: sie sind in der zu der Basis $\langle v_i \rangle$ dualen Basis die zu den v_i dualen Elemente. Auf Grund der Formel (D) bestehen die Basis $[z_i]$ und die Basis $\langle v_i \rangle$ aus denselben Elementen; nun ist die Basis $[z_i]$ zu sich selbst dual (bis auf Vorzeichen); also ist die Basis $[z_i]$ zu der Basis $\langle v_i \rangle$ dual. Man bestätigt auch leicht unmittelbar an der Formel (D), dass das einzige Element der Basis $[z_i]$, das mit dem Element $v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}$ der Basis $\langle v_i \rangle$ den Schnitt $\pm v_0$ ergibt, das Element $z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r}$ ist. Wir wollen diese und noch zwei andere Regeln notieren, die unmittelbar aus (D) folgen:

(1) $$z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r} \cdot (v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}) = \pm v_0,$$

(2) $$z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r} \cdot (v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_r}) = 0,$$ wenn

die Indexmengen $\{i_1, i_2, \cdots, i_r\}$ und $\{j_1, j_2, \cdots, j_r\}$ voneinander verschieden sind,

(3) $$z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r} \cdot (v_{k_1} \circ v_{k_2} \circ \cdots \circ v_{k_s}) = 0,$$ wenn $s < r$.

Um sie zu beweisen, hat man nur die P-Produkte der v_i mit Hilfe von (D) als Schnitt der z_i zu schreiben, und dann die Rechenregeln für $\mathfrak{R}(G)$ anzuwenden.

10. Es seien $x = z_{i_1} \cdot z_{i_2} \cdot \cdots \cdot z_{i_r}$ und $y = z_{j_1} \cdot z_{j_2} \cdot \cdots \cdot z_{j_{l-r}}$ zwei Elemente der Basis $[z_\iota]$, und es sei $x \cdot y = \pm v_0$, d. h. die Elemente $z_{i_1}, \cdots, z_{i_r}, z_{j_1}, \cdots, z_{j_{l-r}}$ sind die Elemente $z_1, z_2, \cdots, z_l$, nur in anderer Reihenfolge. Auf Grund von (D) ist dann

$$x \circ y = v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_{l-r}} \circ v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} ;$$

und dieses Produkt ist gleich ± 1, weil die darin auftretenden v_i bis auf die Reihenfolge mit den Elementen $v_1, v_2, \cdots, v_l$ übereinstimmen. Führt man die elementare Bestimmung des Vorzeichens durch, so erhält man: Aus $x \cdot y = v_0$ folgt $x \circ y = 1$. Daraus folgt leicht allgemein: Sind X, Y zwei beliebige Elemente von $\mathfrak{B}(G)$ mit $d(X) + d(Y) = n$, dann ist die Schnittzahl von X und Y gleich ihrer "Aufspannzahl," d. h. in $X \cdot Y = a \cdot v_0$ und in $X \circ Y = b \cdot 1$ sind die Koeffizienten a und b gleich; diese Gleichheit ist ja anschaulich sehr naheliegend.

Eine letzte Bemerkung ist folgende:

Die Elemente $v_1, \cdots, v_r$ seien irgendwelche, linear unabhängige Elemente von $\mathfrak{B}(G)$; dann sind die Elemente

$$v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_s} \qquad (1 \leqq i_k \leqq r,\ i_1 < i_2 < \cdots < i_s)$$

linear unabhängige Elemente von $\mathfrak{B}(G)$. Zum Beweis hat man nur die $v_1, \cdots, v_r$ in eine Basis $v_1, \cdots, v_l$ von $\mathfrak{B}(G)$ aufzunehmen, und mit dieser Basis die Basis $\langle v_\iota \rangle$ von $\mathfrak{B}(G)$ zu bilden.

Kapitel II

Erste Anwendungen des Aufspannsatzes

Der Aufspannsatz soll jetzt mit den wichtigsten gruppentheoretischen Begriffen in Zusammenhang gebracht werden: mit dem der homomorphen Abbildung und hauptsächlich denen der Untergruppe und der Nebengruppenzerlegung.

1. Homomorphe Abbildung

1. G und G' seien zwei Gruppenmannigfaltigkeiten. Eine Abbildung h von G in G' heisst homomorphe Abbildung, wenn sie

a) eine homomorphe Abbildung im gewöhnlichen gruppentheoretischen Sinn ist, und

b) eine stetige Abbildung der Mannigfaltigkeit G in die Mannigfaltigkeit G' ist.

Die homomorphe Abbildung h bildet den Ring $\mathfrak{P}(G)$ homomorph in den Ring $\mathfrak{P}(G')$ ab; die additive Homomorphie ist klar; die Gleichung

$$h(x \circ y) = h(x) \circ h(y)$$

für zwei beliebige Homologieklassen x, y von G und ihre Bilder $h(x)$, $h(y)$ bei h beweist man etwa folgendermassen:

Man bilde das topologische Produkt $G \times G$ durch die Abbildung $f(p \times q) = h(pq)$ in G' ab. Ist F die Abbildung $F(p \times q) = pq$ von $G \times G$ in G, ist F' die

entsprechende Abbildung für G', und bezeichnet man mit g die Abbildung $g(p \times q) = h(p) \times h(q)$ von $G \times G$ in $G' \times G'$, dann kann man wegen $h(p \cdot q) = h(p) \cdot h(q)$ die Abbildung f sowohl als hF als auch als $F'g$ schreiben. Nun ist $F(x \times y) = x \circ y$, also $hF(x \times y) = h(x \circ y)$. Andererseits ist (nach Kap. I, §1, Nr. 8) $g(x \times y) = h(x) \times h(y)$, und folglich $F'g(x \times y) = h(x) \circ h(y)$. Damit ist die Behauptung bewiesen.

Ist speziell $h(x)$ homolog Null in G', so folgt, dass auch $h(x \circ y)$ homolog Null ist für beliebiges y; denn es ist $0' \circ z' = 0'$ ($0'$ die Nullklasse, z' eine beliebige Homologieklasse von G').

2. Der Kern einer homomorphen Abbildung eines Ringes bzw. einer Gruppe ist die Menge derjenigen Elemente, deren Bild die Null ist; der Kern ist ein Ideal bzw. eine (invariante) Untergruppe.

Wir betrachten die Gruppen $\mathfrak{V}(G)$ und $\mathfrak{V}(G')$ der minimalen Elemente von G und G'. Durch h wird $\mathfrak{V}(G)$ homomorph in $\mathfrak{V}(G')$ abgebildet (nach Kap. I, §1, Nr. 4); $\mathfrak{V}^0(G)$ sei der Kern dieser Abbildung. $\mathfrak{V}(G)$ werde als direkte Summe $\mathfrak{V}^0(G) + \mathfrak{V}^1(G)$ dargestellt; dann wird $\mathfrak{V}^1(G)$ isomorph abgebildet. $\{v_1, \cdots, v_{l^0}\}$ sei eine Basis von $\mathfrak{V}^0$, $\{v_{l^0+1}, \cdots, v_l\}$ eine von $\mathfrak{V}^1$; $\{v_1, \cdots, v_l\}$ ist dann eine Basis von $\mathfrak{V}(G)$. Die Elemente $v_1, \cdots, v_{l^0}$ werden auf Null abgebildet; die Elemente $h(v_{l^0+1}), \cdots, h(v_l)$ sind in G' linear unabhängig.

Wir betrachten die von v_0 und den Produkten

$$v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} \qquad (i_1 < i_2 < \cdots < i_r)$$

gebildete Basis $\langle v_i \rangle$ von $\mathfrak{V}(G)$ (vgl. Satz I a)). Nach der letzten Bemerkung in Nr. 1 wird ein solches Produkt sicher auf 0 abgebildet, wenn auch nur ein Faktor v_{i_k} auf 0 abgebildet wird, wenn also wenigstens ein Index i_k aus der Reihe 1, 2, $\cdots, l^0$ stammt. Andererseits werden diejenigen Elemente $v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_s}$, in denen alle Indizes $> l^0$ sind, auf die Elemente

$$h(v_{j_1} \circ v_{j_2} \circ \cdots \circ v_{j_s}) = h(v_{j_1}) \circ h(v_{j_2}) \circ \cdots \circ h(v_{j_s})$$

abgebildet, und diese Elemente sind in G' linear unabhängig, weil die—nach Kap. I, §1, Nr. 4 minimalen—Elemente $h(v_j)$ $(j = l^0 + 1, \cdots, l)$ linear unabhängig sind (Kap. I, §3, Nr. 10).

Das bedeutet offenbar, dass der Charakter der Abbildung h durch folgenden Satz bestimmt ist:

Satz II. *Ist h eine homomorphe Abbildung von G in G', so ist der Kern der dadurch bewirkten homomorphen Abbildung von $\mathfrak{P}(G)$ in $\mathfrak{P}(G')$ das Ideal von $\mathfrak{P}(G)$, das erzeugt wird von denjenigen minimalen Elementen von $\mathfrak{R}(G)$, deren Bild bei h die Null ist.*

Der Restklassenring von $\mathfrak{P}(G)$ nach diesem Ideal—der ja durch h isomorph abgebildet wird—, ist isomorph dem von den Elementen $v_0, v_{l^0+1}, \cdots, v_l$ erzeugten Teilring von $\mathfrak{P}(G)$.

In jedem von Null verschiedenen Ideal von $\mathfrak{P}(G)$ ist das Element 1 (die Eins von $\mathfrak{R}(G)$) enthalten; denn wegen der Gleichung $v_1 \circ v_2 \circ \cdots \circ v_l = 1$ gibt es zu

jedem Element $x(\neq 0)$ ein y mit $x \circ y = 1$. Daher ist eine direkte Konsequenz des Satzes II das folgende.

KOROLLAR 1. *Das Bild $h(1)$ des Elementes 1 ist dann und nur dann ~ 0 in G', wenn durch h wenigstens ein minimales Element ($\neq 0$) von G auf 0 abgebildet wird. Ist $h(1) \nsim 0$, dann wird der ganze Ring $\mathfrak{P}(G)$ isomorph abgebildet.*

Daraus schliesst man nun weiter:

KOROLLAR 2. *Ist G durch f homomorph in G' abgebildet, und G' durch g homomorph in G'' abgebildet, ist weiter $f(G) \nsim 0$ in G', und $g(G') \nsim 0$ in G'', so ist $gf(G) \nsim 0$ in G''* (G, G' bezeichnen hier die Einselemente von $\mathfrak{R}(G)$, $\mathfrak{R}(G')$).

Denn durch den Homomorphismus gf wird sogar der ganze Ring $\mathfrak{P}(G)$ isomorph abgebildet, weil f und g Isomorphismen sind.

3. In Anwendung dieser Begriffe soll jetzt die Abbildung $q(p) = p^2$ von G in sich, die also jedem Punkt p von G sein Quadrat p^2 zuordnet, untersucht werden.

Bezeichnet man mit f die Abbildung $f(p) = p \times p$ von G in $G \times G$ und mit F die schon betrachtete Abbildung $F(p \times p') = p \cdot p'$, so kann man die Abbildung $q(p)$ als zusammengesetzte Abbildung $Ff(p)$ schreiben. Das Produkt $G \times G$ ist selbst eine Gruppe, als direktes Produkt von G und G; und $f(p) = p \times p$ ist eine homomorphe Abbildung von G in $G \times G$.

Nach Nr. 2 wird man die f-Bilder der minimalen Elemente von G betrachten. Wir behaupten: Für jedes Element v von $\mathfrak{B}(G)$ ist

$$f(v) = (v \times v_0) + (v_0 \times v).$$

Zum Beweis beachten wir, dass $f(v)$ als minimales Element von $\mathfrak{R}(G \times G)$ von der Form $(v^1 \times v_0) + (v_0 \times v^2)$ mit $v^1, v^2 \in \mathfrak{B}(G)$ ist (Kap. I, §1, Nr. 9). Um zu zeigen, dass $v^1 = v^2 = v$ gilt, betrachten wir die Abbildungen $\pi_1(p \times p') = p$ und $\pi_2(p \times p') = p'$ von $G \times G$ in G; die Abbildungen $\pi_1 f$ und $\pi_2 f$ sind dann die Identität von G. Man erkennt sofort: Für jedes Element X der Form $(x \times v_0) + (v_0 \times y)$ ist $\pi_1(X) = x$ und $\pi_2(X) = y$; für jedes z von $\mathfrak{B}(G)$ ist $\pi_1 f(z) = \pi_2 f(z) = z$. Damit hat man die folgenden Gleichungen:

$$v = \pi_1 f(v) = \pi_1((v^1 \times v_0) + (v_0 \times v^2)) = v^1;$$

ebenso zeigt man: $v = v^2$, womit die obige Behauptung bewiesen ist.

Sei jetzt $\{v_1, v_2, \cdots, v_l\}$ eine Basis von $\mathfrak{B}(G)$. Nach dem eben Bewiesenen ist dann

$$f(v_i) = (v_i \times v_0) + (v_0 \times v_i),$$

also

$$q(v_i) = Ff(v_i) = F((v_i \times v_0) + (v_0 \times v_i)) = v_i + v_i = 2v_i\,.$$

Die Bilder $f(v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r})$ der Elemente der Basis $\langle v_i \rangle$ bestimmt man jetzt daraus, dass f homomorph ist. Man hat dabei zu beachten, dass für die Pontrjagin'sche Multiplikation in $G \times G$, wie man leicht sieht, gilt:

$$(x \times y) \circ (x' \times y') = (-1)^{d(y)d(x')}(x \circ x') \times (y \circ y').$$

Damit läuft die Rechnung so:

$$\begin{aligned} f(v_i \circ v_k) &= f(v_i) \circ f(v_k) \\ &= ((v_i \times v_0) + (v_0 \times v_i)) \circ ((v_k \times v_0) + (v_0 \times v_k)) \\ &= ((v_i \circ v_k) \times v_0) + (v_i \times v_k) - (v_k \times v_i) + (v_0 \times (v_i \circ v_k)); \end{aligned}$$

also

$$\begin{aligned} q(v_i \circ v_k) &= Ff(v_i \circ v_k) \\ &= F((v_i \circ v_k) \times v_0) + F(v_i \times v_k) - F(v_k \times v_i) + F(v_0 \times (v_i \circ v_k)) \\ &= (v_i \circ v_k) + (v_i \circ v_k) - (v_k \circ v_i) + (v_i \circ v_k); \end{aligned}$$

und wegen $v_i \circ v_k = -v_k \circ v_i$ erhält man endlich:

$$q(v_i \circ v_k) = 4 v_i \circ v_k .$$

Entsprechend verfährt man im allgemeinen Falle, bei der Berechnung von $q(v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}) = Ff(v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r})$. Man geht aus von der Gleichung

$$f(v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}) = f(v_{i_1}) \circ f(v_{i_2}) \circ \cdots \circ f(v_{i_r}),$$

setzt $f(v_{i_k}) = (v_{i_k} \times v_0) + (v_0 \times v_{i_k})$ ein, und multipliziert aus. Man erhält 2^r Summanden, von denen jeder das Element $\pm v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}$ als Bild bei F hat—und man stellt fest, dass immer das positive Vorzeichen gilt. Damit hat man den Satz:

Der Homologietyp der Abbildung $q(p) = p^2$ von G in sich ist bestimmt durch die folgenden Gleichungen, in denen die q-Bilder der Elemente der Basis $\langle v_i \rangle$ von $\mathfrak{B}(G)$ angegeben sind:

$$q(v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}) = 2^r v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} \qquad (i_1 < i_2 < \cdots < i_r)$$

(und natürlich $q(v_0) = v_0$).

Darin ist, wegen der Gleichung $1 = v_1 \circ v_2 \circ \cdots \circ v_l$, enthalten:[5] Die Abbildung $q(p) = p^2$ von G in sich hat den Grad 2^l.

2. Untergruppen und Nebengruppenzerlegung

1. G sei wie immer eine Gruppenmannigfaltigkeit. Eine Teilmenge U heisst eine Untergruppe, wenn sie

a) eine Untergruppe von G im gewöhnlichen gruppentheoretischen Sinn ist, und

b) eine abgeschlossene Teilmenge von G ist.

Dazu bemerken wir folgendes: Es ist bekannt, dass jede geschlossene Gruppenmannigfaltigkeit eine Lie'sche Gruppe, also jedenfalls einige Male differenzierbar ist. Ueber das Verhalten einer Untergruppe U ist bekannt [17]: Man kann in einer hinreichend kleinen, dem Euklidischen R_n ($n = d(G)$) homöomorphen,

[5] Vgl.: H. Hopf: *Über den Rang geschlossener Liescher Gruppen* [Comm. Math. Helv 13. (1940/41)], Satz I für $k = 2$.

Umgebung V des Einheitspunktes e der Gruppe G ein solches Koordinatensystem einführen, dass diejenigen Punkte von U, die zu V gehören, die also den Durchschnitt $U \cdot V$ bilden, eine gewisse Ebene durch e erfüllen. U ist also im kleinen euklidisch. Beachtet man noch, dass U als abgeschlossene Teilmenge von G kompakt ist, so sieht man leicht: Ist die Untergruppe U zusammenhängend, dann ist sie eine differenzierbare und differenzierbar in G eingelagerte, geschlossene, Gruppenmannigfaltigkeit; ist U nicht zusammenhängend, dann ist die Komponente U', die den Punkt e enthält, selbst eine zusammenhängende Untergruppe von G, also eine Gruppenmannigfaltigkeit in G, und U besteht aus endlich viel Komponenten, die Nebengruppen von U' sind.

Deshalb bedeutet es kaum eine Einschränkung, wenn wir im folgenden die Forderung b) ersetzen durch die Forderung

b′) U ist eine (differenzierbar) in G eingelagerte geschlossene Mannigfaltigkeit (also zusammenhängend).

Da U eine Gruppenmannigfaltigkeit ist, haben für sie die Definitionen und Sätze des Kap. I Gültigkeit.

2. G sei eine Gruppe, U eine Untergruppe. Die Gruppenmannigfaltigkeit U ist dadurch, dass sie eine Untergruppe von G ist, auf natürliche Weise homomorph, sogar isomorph, in G abgebildet; man hat dazu die Punkte von U, die ja auch Punkte von G sind, eben als Punkte von G aufzufassen. Das Bild einer Homologieklasse u von U bei dieser Abbildung ist u, als Homologieklasse von G betrachtet; genauer: das Bild von u ist die Homologieklasse von G, in der die Zyklen aus u, die ja auch Zyklen in G sind, liegen. Die Homomorphie besteht darin, dass für die Homologieelemente von U die Pontrjagin'sche Multiplikation in U übereinstimmt mit der in G.

Man kann also die Sätze des §1 anwenden. Wir stellen die Frage: Welche Elemente von $\mathfrak{B}(U)$ sind homolog 0 in G? Nach §1 hängt das davon ab, welche *minimalen* Elemente von $\mathfrak{R}(U)$ in G homolog 0 sind. Aus dem Satz II folgt der

Satz III. *Die Gesamtheit der Homologieelemente der Untergruppe U, die in der Gruppe G homolog* 0 *sind, ist das Ideal von $\mathfrak{P}(U)$, das erzeugt wird von denjenigen minimalen Elementen von $\mathfrak{R}(U)$, die in G homolog* 0 *sind.*

Das sind also die Elemente $\sum w_h \circ v_h$, wo die w_h beliebige Elemente von $\mathfrak{P}(U)$ sind, und die v_h Elemente von $\mathfrak{B}(U)$ bedeuten, die in G homolog 0 sind.

Daraus entnimmt man sofort das

Korollar 1. *Eine Untergruppe U ist dann und nur dann ~ 0 in der Gruppe G, wenn wenigstens ein von Null verschiedenes minimales Element aus $\mathfrak{R}(U)$ in G homolog* 0 *ist.*

Weiter gilt:

Korollar 2. *Ist U nicht homolog* 0 *in G, dann ist kein von Null verschiedenes Element von $\mathfrak{B}(U)$ homolog* 0 *in G.*

Daraus wieder folgt sofort:

Korollar 3. *Ist G'' eine Untergruppe von G', die nicht homolog* 0 *ist in G', und ist G' eine Untergruppe von G, die nicht homolog* 0 *ist in G, dann ist G'' nicht homolog* 0 *in G.*

Dann ist sogar kein Element von $\mathfrak{B}(G'')$ homolog 0 in G.

Diesen Satz verallgemeinert man zu folgendem:

Ist $G_1 \supset G_2 \supset G_3 \supset \cdots \supset G_r$ eine absteigende Kette von Untergruppen, und ist $G_{i+1} \nsim 0$ in G_i (für $i = 1, 2, \cdots, r-1$), dann ist $G_r \nsim 0$ in G_1.

Man kann die Ueberlegungen, die zum Satz III führen, etwas anders fassen: $\mathfrak{B}^0(U)$ sei die Gruppe derjenigen Elemente von $\mathfrak{B}(U)$, die ~ 0 in G sind. Man stelle $\mathfrak{B}(U)$ als direkte Summe $\mathfrak{B}^0(U) + \mathfrak{B}^1(U)$ dar. Die Elemente $v_1, v_2, \cdots, v_r$ einer Basis von $\mathfrak{B}^1(U)$ sind dann in G linear unabhängig; man kann sie sowohl in eine Basis von $\mathfrak{B}(U)$ als auch in eine Basis von $\mathfrak{B}(G)$ aufnehmen. Die Produkte $v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_s}$ ($i_1 < i_2 < \cdots < i_r$, $1 \leqq i_k \leqq r$), die man sowohl als Elemente von $\mathfrak{P}(U)$ als auch als Elemente von $\mathfrak{P}(G)$ auffassen kann, sind nach dem Aufspannsatz sowohl in $\mathfrak{P}(U)$ als auch in $\mathfrak{P}(G)$ linear unabhängig; sie bilden eine Basis der Gruppe derjenigen Homologieklassen von G, in denen Homologieklassen von U liegen; die Elemente von $\mathfrak{B}(U)$, die ~ 0 in G sind, sind die Elemente $\sum w_h \circ v_h$ mit $w_h \subset \mathfrak{B}(U)$, $v_h \subset \mathfrak{B}^0(U)$.

Gleichbedeutend damit, dass $U \nsim 0$ in G ist, ist: $\mathfrak{B}^0(U)$ enthält nur die Null.

3. Ist G eine Gruppe, U eine Untergruppe, so zerfällt G in die (etwa linksseitigen) Nebengruppen, was man durch die Gleichung

$$G = U + a_1 \cdot U + a_2 \cdot U + \cdots$$

ausdrückt, in der die a_i gewisse Elemente von G sind. (Wir gebrauchen die übliche Schreibweise auch hier, wo die Anzahl der Nebengruppen kontinuierlich ist.)

Die Wichtigkeit dieser Nebengruppenzerlegung für topologische Gruppen G besteht darin, dass man die Nebengruppen als Punkte eines topologischen Raumes, des Nebengruppenraumes, auffassen kann, dessen Topologie eng mit der von G verknüpft ist. Dieser Raum, der mit G/U oder mit W ("Wirkungsraum," vgl. Nr. 5) bezeichnet wird, ist also folgendermassen definiert:

Die Punkte von W sind die Nebengruppen aU von U; eine Umgebung eines Punktes aU besteht aus den Nebengruppen $a' \cdot U$, wo der Punkt a' eine Umgebung von a in G durchläuft.

Auf Grund der Bemerkungen in Nr. 1 ist leicht zu sehen, dass auch der Nebengruppenraum eine differenzierbare Mannigfaltigkeit ist. Man führt dazu in einer (hinreichend kleinen) Umgebung des beliebigen Punktes a von G ein Koordinatensystem so ein, dass die Nebengruppe aU durch eine r-dimensionale Ebene ($r = d(U)$) durch a dargestellt wird; man stellt leicht fest, dass die Nebengruppen, die eine Umgebung von aU in W bilden, eineindeutig den Punkten der zu aU senkrechten, $(n - r)$-dimensionalen Ebene durch a entsprechen ($n = d(G)$. Damit hat man in der Umgebung des Punktes aU von W ein Euklidisches Koordinatensystem eingeführt. Die Differenzierbarkeit zeigt man ebenso.

Wir könne und wollen also im folgenden annehmen, dass der Nebengruppen-

raum W eine differenzierbare, also triangulierbare [18], geschlossene Mannigfaltigkeit ist.

4. Die Zerlegung von G in die Nebengruppen aU ist eine Faserung mit der Faser U; dabei versteht man unter der Faserung einer Mannigfaltigkeit M mit der Faser (-Mannigfaltigkeit) F folgendes [10]: M ist zerlegt in Mannigfaltigkeiten, die alle einer festen Mannigfaltigkeit F homöomorph sind, und die "Fasern" von M heissen; durch jeden Punkt von M geht also eine und nur eine Faser; und eine Umgebung einer beliebigen Faser ist fasertreu homöomorph dem topologischen Produkt $E \times F$, wo E ein Element (Vollkugel) der Dimension $d(E) = d(M) - d(F)$ ist.

"Fasertreu homöomorph" heisst dabei: jede Faser der Umgebung ist topologisch auf eine der Mannigfaltigkeiten $p \times F$ abgebildet (p ein Punkt von E)—verschiedene Fasern natürlich auf verschiedene $p \times F$.

Fasst man die Fasern als Punkte eines Raumes, mit natürlichem Umgebungsbegriff, auf (bildet man den sog. Zerlegungsraum der Zerlegung von M in die Fasern), so erhält man den Faser- oder Basisraum B, der offenbar eine Mannigfaltigkeit ist; der Faserraum wird auch, in einer naheliegenden Schreibweise, mit M/F bezeichnet.

Wir sprechen im folgenden von Faserungen auch bei Teilmengen einer gefaserten Mannigfaltigkeit, die ganz aus Fasern bestehen, d. h. zu einer beliebigen Faser entweder fremd sind oder sie ganz enthalten; und ist K ein beliebiger Komplex, so nennen wir das topologische Produkt $K \times F$ gefasert—die Fasern sind natürlich die Mengen $p \times F$, wo p ein beliebiger Punkt von K ist. Es ist dann klar, was man unter einer fasertreuen (topologischen) Abbildung zu verstehen hat.

Ordnet man jedem Punkt einer gefaserten Mannigfaltigkeit M die Faser zu, auf der er liegt, so erhält man eine stetige Abbildung von M auf den Faserraum B, die Projektion P. Das Urbild $P^{-1}(T)$ einer Teilmenge T von B ist die (gefaserte) Menge der Punkte der "über T stehenden" Fasern. Für jede hinreichend kleine Umgebung V eines beliebigen Punktes von B ist $P^{-1}(V)$ fasertreu homöomorph mit $V \times F$; M ist im kleinen ein topologisches Produkt. Ueber dieses lokale Zerfallen in ein topologisches Produkt gilt der folgende Satz von Feldbau [19]:

Ist E ein Element (Vollkugel) im Faserraum B, dann ist die Urbildmenge $P^{-1}(E)$ dem Produkt $E \times F$ fasertreu homöomorph.

Nach Nr. 3 ist leicht zu sehen, dass die Zerlegung einer Gruppe G in die Nebengruppen aU einer Untergruppe U eine Faserung von G mit der Faser U ist. Man erkennt nämlich sofort: Durchläuft p die in Nr. 3 genannte, zu aU senkrechte Ebene, die E heissen möge, in der Umgebung von a, und durchläuft q die Untergruppe U, dann ist $f(p \times q) = p \cdot q$ eine fasertreue topologische Abbildung von $E \times U$ auf die Umgebung von U (pq ist das Produkt von p und q in G).

Der Faserraum dieser Faserung ist der Nebengruppen- oder Wirkungsraum W. Die Projektion P besteht darin, dass man dem Punkt p von G die Nebengruppe pU (als Punkt von W) zuordnet.

5. Eine Mannigfaltigkeit W heisst Wirkungsraum der Gruppenmannigfaltigkeit G ("espace homogène") [9], wenn folgendes erfüllt ist:

Jedem Punkt p von G ist eine topologische Transformation T_p von W auf sich zugeordnet; die Transformationen T_p hängen stetig von dem Punkt p ab; es gilt: $T_p \cdot T_q = T_{pq}$; und die aus den T_p bestehende Transformationsgruppe von W ist transitiv, d. h. jeder Punkt von W (es genügt: ein bestimmter Punkt von W) kann (durch geeignete Transformationen T_p) in jeden Punkt von W übergeführt werden.

a sei ein Punkt von W; diejenigen Punkte p von G, für die $T_p(a) = a$ gilt, d. h. deren T_p den Punkt a als Fixpunkt hat, bilden eine Untergruppe von G, die "Isotropiegruppe" von W. (Setzt man $W = R_n$ und nimmt als G die Gruppe der Bewegungen des R_n, so ist die Isotropiegruppe die Gruppe der Rotationen um einen festen Punkt.) Die zu verschiedenen Punkten a von W gehörigen Isotropiegruppen sind in G konjugiert, sodass man von *der* Isotropiegruppe U sprechen kann.

Die Punkte p von G, deren T_p den Punkt a in den gleichen Punkt b transformieren, bilden eine linke Nebengruppe von U. Diese eineindeutige Zuordnung der Punkte von W und der linken Nebengruppen von U ist eine Homöomorphie zwischen W und dem Nebengruppenraum G/U.

Die Transformationen T_p können also auch als Transformationen des Nebengruppenraumes G/U aufgefasst werden; sie haben da eine einfache Bedeutung: Die Transformation T_p besteht nämlich darin, dass man jeder Nebengruppe qU die Nebengruppe pqU zuordnet.

Ein Wirkungsraum W ist also identisch mit dem zur Isotropiegruppe U von W gehörigen Nebengruppenraum G/U (bei der Definition des Nebengruppenraums muss man dann mehrkomponentige Untergruppen (vgl. Nr. 1) U zulassen).

Diese Definition des Wirkungsraumes ist etwas allgemeiner als die übliche: wir lassen zu, dass T_p für gewisse p die Identität von W ist (p gehört dann zu U).

Ist U zusammenhängend, dann ist W orientierbar (aber nicht nur dann): Man stelle (nach Nr. 4) die Umgebung einer Nebengruppe pU als topologisches Produkt $E \times U$ dar. Die Orientierung von G und die von U bestimmen dann eindeutig eine Orientierung von E, und diese überträgt man mittels der Projektion P in den Wirkungsraum W. (Entsprechend für Faserungen M/F.)

6. Das Problem, das sich jetzt erhebt, ist: Zusammenhänge zwischen G, U, W, insbesondere zwischen ihren Homologieeigenschaften, zu finden. Es handelt sich z. B. um die Aufgabe, aus den Homologieeigenschaften von G und U die von W zu bestimmen, wobei man natürlich die Lage von U in G berücksichtigen muss. Zur Lösung dieser Aufgabe werden in Kap. IV einige Beiträge geliefert. Noch naheliegender ist die Fragestellung: was kann man bei gegebenem G und W über U aussagen?, also die Frage nach der Isotropiegruppe U eines vorgelegten Wirkungsraumes W. Diese Frage wird im Kap. III für den speziellen Fall $W = S_n$ gelöst; als Anwendung werden die Homologieringe der Gruppen A_n, B_n, C_n, D_n bestimmt.

Kapitel III

Die Sphären als Wirkungsräume

In diesem Kapitel wird der Spezialfall untersucht, dass der zu der Gruppe G und der Untergruppe U gehörige Wirkungsraum W der m-dimensionalen Sphäre S_m homöomorph ist. Dieser Fall tritt auf bei den Gruppen der bekannten vier Klassen A_n, B_n, C_n, D_n; die Homologieringe dieser Gruppen lassen sich damit vollständig bestimmen.

1. $G/U = S_m$

1. G sei eine Gruppe; der Rang von G (Kap. I, §2, Nr. 3) sei l; U sei eine Untergruppe von G. Wir beweisen den folgenden Satz:

Satz IV. *Der zu der Untergruppe U der Gruppe G gehörige Wirkungsraum $G/U = W$ sei der m-dimensionalen Sphäre S_m homöomorph; dann gilt:*

a) *ist m ungerade, so ist $\mathfrak{R}(G)$ isomorph mit $\mathfrak{R}(U \times S_m)$, und U ist $\nsim 0$ in G,*

b) *ist m gerade, so ist $\mathfrak{R}(U)$ isomorph mit $\mathfrak{R}(\Pi \times S_{m-1})$, wo Π ein topologisches Produkt von $l - 1$ Sphären ungerader Dimensionen ist, und $\mathfrak{R}(G)$ ist isomorph mit $\mathfrak{R}(\Pi \times S_{2m-1})$; und U ist ~ 0 in G.*

Bemerkung: für $l = 1$ bedeutet b): $\mathfrak{R}(U) \approx \mathfrak{R}(S_{m-1})$, und $\mathfrak{R}(G) \approx \mathfrak{R}(S_{2m-1})$. (Dann ist übrigens, was aber für das folgende unwesentlich ist, notwendigerweise $m = 2$.)

Für die Ränge $l(G)$, $l(U)$ und die charakteristischen Zahlen $l_i(G)$, $l_i(U)$ (Kap. I, §2, Nr. 3) bedeutet dieser Satz das Bestehen der folgenden Gleichungen:

a) m ungerade:

$$l(G) = l(U) + 1,$$
$$l_m(G) = l_m(U) + 1,$$
$$l_i(G) = l_i(U) \qquad \text{für } i \neq m,$$

b) m gerade:

$$l(G) = l(U)$$
$$l_{m-1}(G) = l_{m-1}(U) - 1,$$
$$l_{2m-1}(G) = l_{2m-1}(U) + 1,$$
$$l_i(G) = l_i(U) \qquad \text{für } i \neq m-1, \neq 2m-1.$$

Der Beweis beruht auf der Betrachtung der Faserung von G in die Nebengruppen aU von U.

2. Wir erinnern zunächst an einige Begriffe:

Ein endliches, euklidisches Polyeder ist die Menge der Punkte, die zu einem endlichen, euklidischen, simplizialen (allgemeiner: Zellen-) Komplex gehören. Ein endliches, krummes Polyeder (kurz: Polyeder) P in einer Mannigfaltigkeit

M ist das topologische Bild in M eines endlichen, euklidischen Polyeders [20]. Ein stetiger Zyklus ist das stetige Bild eines, in einem euklidischen R_n liegenden, simplizialen "Parameter"-Zyklus. Ein stetiger Zyklus z in einem Polyeder P bestimmt eine Homologieklasse von P, die man eben durch z repräsentieren kann [21]. Im folgenden verstehen wir unter Zyklen immer stetige Zyklen.

Wir betrachten eine beliebige Gruppe G. Im Kap. I, §3, Nr. 1 ist mit Hilfe der Abbildung $F(p \times q) = pq$ von $G \times G$ in G das Pontrjagin'sche Produkt $x \circ y$ zweier Homologieklassen x und y von G definiert worden. Diese Definition lässt sich unmittelbar erweitern zu der des Pontrjagin'schen Produktes $z_1 \circ z_2$ zweier stetiger Zyklen z_1, z_2 in G: man bilde den stetigen Zyklus $z_1 \times z_2$ in $G \times G$, und definiere $z_1 \circ z_2$ durch

$$z_1 \circ z_2 = F(z_1 \times z_2);$$

$z_1 \circ z_2$ ist ebenfalls ein stetiger Zyklus in G.

3. Jetzt sei gemäss der Voraussetzung von Satz IV der zur Untergruppe U der Gruppe G gehörige Nebengruppenraum G/U der S_m homöomorph. P bezeichnet die Projektion von G auf $G/U = S_m$.

Wir zerlegen die S_m durch eine $(m - 1)$-dimensionale Aequatorsphäre S_{m-1} in zwei Elemente (Vollkugeln) E^1 und E^2, die "nördliche und südliche Halbkugel." Es ist also $E^1 + E^2 = S_m$ und $E^1 \cdot E^2 = S_{m-1}$. Das Urbild $P^{-1}(E^1)$ von E^1 heisse G', das Urbild $P^{-1}(E^2)$ von E^2 heisse G''. Es ist $G = G' + G''$; d. h., G wird von G' und G'' überdeckt. Der Durchschnitt $G' \cdot G''$, der H heisse, ist das Urbild $P^{-1}(S_{m-1})$ von S_{m-1}.

Wir wenden jetzt den Satz von Feldbau an (Kap. II, §2, Nr. 4). Weil E^1 ein Element ist, ist danach die Menge G', die ja ganz aus Nebengruppen von U besteht, dem topologischen Produkt $E^1 \times U$ fasertreu homöomorph. Dabei ist nur benutzt, dass eine Faserung von G in die Nebengruppen von U vorliegt. Indem man nun die Gruppeneigenschaft stärker ausnützt, kann man die Darstellung von G' als Produkt $E^1 \times U$ in einer speziellen Weise wählen. Man betrachte dazu für einen beliebig, aber fest gewählten Punkt q von U das in $E^1 \times U$ liegende Element $E^1 \times q$. Ihm entspricht infolge der Homöomorphie von $E^1 \times U$ und G' ein Element E' in G'. Aus der Definition folgt unmittelbar, dass E' mit jeder zu G' gehörenden Faser von G (Nebengruppe von U) genau einen, und mit jeder nicht zu G' gehörenden Faser keinen Punkt gemeinsam hat und dass E' durch P topologisch auf E^1 abgebildet wird. Man betrachte nun das in $G \times G$ liegende Polyeder $E' \times U$. Bei der Abbildung F wird dieses Polyeder topologisch-fasertreu auf G' abgebildet. Denn ist p ein beliebiger Punkt von E', so wird die Teilmenge $p \times U$ von $E' \times U$ topologisch auf die, in G' enthaltene, Nebengruppe pU abgebildet (das ist geradezu die Definition der Nebengruppe); sind p_1, p_2 zwei verschiedene Punkte von E', so sind, wie eben bemerkt, die Nebengruppen $p_1 \cdot U$ und $p_2 \cdot U$, auf die die Mengen $p_1 \times U$ und $p_2 \times U$ abgebildet werden, verschieden; und man erhält so auch alle zu G' gehörenden Nebengruppen. Die ursprüngliche Homöomorphie zwischen

$E^1 \times U$ und G' ist also jetzt ersetzt durch die durch F vermittelte Homöomorphie zwischen $E' \times U$ und G'. Entsprechend lässt sich G'' darstellen als topologisches F-Bild eines in $G \times G$ liegenden Produktes $E'' \times U$, wo E'' ein Element in G ist, das durch P topologisch auf E^2 abgebildet wird. G' und G'' sind also krumme Polyeder in G; sie sind sogar berandete Mannigfaltigkeiten.

Die Randsphäre von E' heisse S', die von E'' heisse S''. Beide Sphären werden durch P topologisch auf S_{m-1} abgebildet. Dann sieht man: Bei der topologisch-fasertreuen Abbildung F von $E' \times U$ auf G' erscheint als Bild des Randes $S' \times U$ von $E' \times U$ die Menge $H = P^{-1}(S_{m-1}) = G' \cdot G''$. Denn $S' \times U$ wird bei F topologisch auf die Menge der Punkte derjenigen Nebengruppen von U abgebildet, die einen Punkt mit S' gemeinsam haben, deren Projektion, d. h. deren Bild bei P, also zu S_{m-1} gehört. Als topologisches Bild der Mannigfaltigkeit $S' \times U$ ist H eine Mannigfaltigkeit in G.

Ebenso stellt man fest, dass H topologisch-fasertreues F-Bild des Randes $S'' \times U$ von $E'' \times U$ ist.

4. Den orientierten Grundzyklus der Sphäre S' bezeichnen wir auch mit S'; p' sei ein einfach gezählter Punkt von S'; die Zyklen $u_1, \cdots, u_r$ mögen eine Basis von $\mathfrak{B}(U)$ bilden. Dann bilden die Zyklen $p' \times u_i$ und $S' \times u_i$ $(i = 1, \cdots, r)$ eine Basis von $\mathfrak{B}(S' \times U)$; für das topologische Bild $H = F(S' \times U)$ bedeutet das: die (nach Nr. 2 gebildeten) Zyklen

$$p' \circ u_i, \qquad S' \circ u_i \qquad\qquad (i = 1, \cdots, r)$$

bilden eine Basis von $\mathfrak{B}(H)$; diese Basis heisse $\mathfrak{B}_1$. Ebenso folgt aus der Tatsache, dass H topologisches F-Bild von $S'' \times U$ ist, dass die Zyklen

$$p'' \circ u_i, \qquad S'' \circ u_i \qquad\qquad (i = 1, \cdots, r)$$

eine Basis von $\mathfrak{B}(H)$ bilden; p'', S'' sind analog wie p', S' definiert.

Man sieht nun leicht ein, dass man die beiden Punkte p' und p'' und ebenso den einfach gezählten Punkt p_0 von $\mathfrak{B}(U)$ durch den Einheitspunkt e von G repräsentieren kann. (Man kann zunächst S_{m-1} so wählen, dass U in $P^{-1}(S_{m-1}) = H$ liegt, und kann dann die Elemente E' und E'' so bestimmen, dass sie den Punkt e in ihrem Rand enthalten.)

Man betrachte die Untergruppe von $\mathfrak{B}(H)$, die von den Elementen $p' \circ u$ mit beliebigem u aus $\mathfrak{B}(U)$ gebildet wird; sie ist mit $\mathfrak{B}(U)$ isomorph (entsprechend der Tatsache, dass die Gruppe der Elemente $p' \times u$ von $\mathfrak{B}(S' \times U)$ mit $\mathfrak{B}(U)$ isomorph ist). Wegen der speziellen Wahl von p' ist

$$p' \circ u = e \circ u = u;$$

das bedeutet: man kann die Elemente $p' \circ u$ dieser Untergruppe mit den Elementen u von $\mathfrak{B}(U)$ identifizieren, man kann also $\mathfrak{B}(U)$ als Untergruppe von $\mathfrak{B}(H)$ auffassen. Die Elemente $S' \circ u$ (und ebenso die Elemente $S'' \circ u$) gehören nicht zu dieser Untergruppe. Wir bemerken schliesslich noch, dass wir, wegen $p_0 = e$, das Element $S' \circ p_0$ von $\mathfrak{B}(H)$ mit der Sphäre S' identifizieren können; ebenso identifizieren wir $S'' \circ p_0$ mit S''.

5. Wir stellen jetzt die Frage: Welche Elemente von $\mathfrak{B}(U)$ sind homolog 0 in G? Um sie zu beantworten, verallgemeinern wir sie zu der Frage: Welche Elemente von $\mathfrak{B}(H)$ sind homolog 0 in G?

Wir gehen aus von der Tatsache, dass G überdeckt ist mit den Polyedern G' und G'', deren Durchschnitt eben das Polyeder H ist. Daher lässt sich ein elementarer Satz aus dem Kreis der "Additionssätze der kombinatorischen Topologie" [22] anwenden, der besagt:

Die Homologieklasse z von H ist dann und nur dann homolog 0 in G, wenn sie sich in H darstellen lässt als Summe $z_1 + z_2$ zweier Homologieklassen z_1, z_2 von H, wobei z_1 in G' und z_2 in G'' homolog 0 ist.

(Im rein kombinatorischen Fall verläuft der Beweis dieses Satzes folgendermassen: der in H liegende Zyklus z sei ~ 0 in G, also $z = \dot{C}$. (Mit $\dot{C}$ bezeichnen wir den Rand des Komplexes C.) Unter C_1 verstehe man den Teilkomplex von C, dessen Simplexe in G' liegen, unter C_2 den von den übrigen Simplexen von C gebildeten Teilkomplex von C; es ist also $C = C_1 + C_2$. Dann ist $z = \dot{C} = \dot{C}_1 + \dot{C}_2$, und $\dot{C}_1$, bzw. $\dot{C}_2$, ist ein Zyklus in H, der ~ 0 in G', bzw. G'', ist. Im vorliegenden Fall krummer Polyeder muss man die vorkommenden Zyklen und Komplexe als stetige Zyklen und stetige Komplexe annehmen und noch einige einfache Approximationen vornehmen.)

6. Die Zyklen von H, die in G' homolog 0 sind, sind nun genau die Zyklen $S' \circ u$, wo u ein beliebiges Element von $\mathfrak{B}(U)$ ist, entsprechend der Tatsache, dass in $E' \times U$ genau die Homologieklassen $S' \times u$ des Randes $S' \times U$ von $E' \times U$ homolog 0 sind. Analog sind genau die Elemente $S'' \circ u$ von $\mathfrak{B}(H)$ in $G'' = F(E'' \times U)$ homolog 0. Die Elemente R von $\mathfrak{B}(H)$, die in G homolog 0 sind, sind also nach dem genannten Additionssatz die Elemente

$$R = S' \circ u_1 + S'' \circ u_2 \tag{1}$$

mit beliebigem u_1 und u_2 aus $\mathfrak{B}(U)$.

Um eine Uebersicht über sie zu bekommen, schreiben wir sie in der Basis $\mathfrak{B}_1$ von $\mathfrak{B}(H)$. Dazu stellen wir zunächst einmal S'' in dieser Basis dar:

$$S'' \sim S' \circ u' + p' \circ u \qquad \text{in } H.$$

Aus Dimensionsgründen ist u' gleich cp_0 mit einem gewissen Koeffizienten c. Wegen der in Nr. 4 getroffenen speziellen Wahl von p_0 und p' hat man also:

$$S'' \sim cS' + u \qquad \text{in } H.$$

S' und S'' sind als Sphären minimale Elemente von $\mathfrak{R}(H)$ (vgl. Kap. I, §1, Nr. 4); da die minimalen Elemente eine Gruppe bilden, ist auch u ein minimales Element von $\mathfrak{R}(H)$. Aus dem Zusammenhang zwischen den minimalen Elementen eines topologischen Produktes mit den minimalen Elementen seiner Faktoren (Kap. I, §1, Nr. 9) folgt, dass u sogar ein minimales Element von $\mathfrak{R}(U)$ ist, das wir mit $\bar{v}$ bezeichnen wollen. Es ist also schliesslich:

(2) $$S'' \sim cS' + \bar{v} \qquad \text{in } H.$$

Dabei ist $\bar{v}$ eventuell das Nullelement.

(2) in (1) eingesetzt, ergibt die folgende Gleichung:

$$R \sim S' \circ (u_1 + cu_2) + \bar{v} \circ u_2 \qquad \text{in } H.$$

(Dabei ist benützt, dass für einen Zyklus z in H und einen Zyklus u in U das Produkt $z \circ u$ in H liegt, und dass für zwei Zyklen z_1, z_2 in H und einen Zyklus u in U gilt:

$$(z_1 + z_2) \circ u \sim z_1 \circ u + z_2 \circ u \qquad \text{in } H.$$

Um das zu beweisen, beachte man, dass H aus linken Nebengruppen von U besteht, und dass folglich die Teilmenge $H \times U$ von $G \times G$ durch F in H abgebildet wird. In $H \times U$ ist nun $(z_1 + z_2) \times u$ homolog $(z_1 \times u) + (z_2 \times u)$; übt man darauf F aus, so erhält man die behauptete Homologie.)

Damit hat man schliesslich:

Die Gesamtheit der Elemente von $\mathfrak{B}(H)$, die in G homolog 0 sind, wird gebildet von den Elementen

(3) $$S' \circ u + \bar{v} \circ u_2,$$

wo u, u_2 beliebige Elemente von $\mathfrak{B}(U)$ sind.

Unter den Elementen (3) sind nur die mit $u = 0$ in der Gruppe $\mathfrak{B}(U)$ enthalten; also folgt:

Die Gesamtheit der Elemente von $\mathfrak{B}(U)$, die in G homolog 0 sind, wird gebildet von den Elementen

$$\bar{v} \circ u$$

mit beliebigem u aus $\mathfrak{B}(U)$, sie ist also das von $\bar{v} = S'' - cS'$ erzeugte Ideal von $\mathfrak{P}(U)$.

7. Nach Satz III (Kap. II, §2, Nr. 2) ist daher $\bar{v}$ das einzige minimale Element von $\mathfrak{R}(U)$, das homolog 0 in G ist; genauer: die Gruppe $\mathfrak{B}^0(U)$ derjenigen minimalen Elemente von $\mathfrak{R}(U)$, die in G homolog 0 sind, besteht aus den rationalen Vielfachen von $\bar{v}$.

Aus der Definition von $\bar{v}$ folgt: $d(\bar{v}) = d(S') = d(S'') = m - 1$. Wir unterscheiden jetzt die Fälle $\bar{v} = 0$ und $\bar{v} \neq 0$. Der Kürze halber wird der Rang von $\mathfrak{B}(G)$ mit l und der Rang von $\mathfrak{B}(U)$ mit l' bezeichnet.

a) Es sei $\bar{v} = 0$. Dann ist also (ausser 0) kein minimales Element von $\mathfrak{R}(U) \sim 0$ in G; also ist nach Kap. II, §2, Nr. 2, Korollar 1 die Untergruppe $U \not\sim 0$ in G, und nach Korollar 2 kann man $\mathfrak{B}(U)$ als Untergruppe von $\mathfrak{B}(G)$ auffassen. Eine Basis $\{v_1, v_2, \cdots, v_{l'}\}$ von $\mathfrak{B}(U)$ kann man aufnehmen in eine Basis $\{v_1, v_2, \cdots, v_{l'}, v_{l'+1}, \cdots, v_l\}$ von $\mathfrak{B}(G)$. Wegen $\sum_1^{l'} d(v_i) = d(U)$, $\sum_1^{l} d(v_i) = d(G)$ und $d(G) = d(U) + m$ gilt dabei:

$$d(v_{l'+1}) + d(v_{l'+2}) + \cdots + d(v_l) = m.$$

Nun ist aber jede Homologieklasse von G, deren Dimension kleiner als m ist, in G einer Homologieklasse von U homolog. Denn aus der Tatsache, dass E^i ein m-dimensionales Element ist ($i = 1, 2$), schliesst man ohne weiteres, dass man jeden höchstens $(m - 1)$-dimensionalen Zyklus von G auf den gemeinsamen Rand H von G' und G'' deformieren kann. Aber die Homologieelemente von $\mathfrak{B}(H)$, die $\nsim 0$ in G sind, sind in G den Elementen der Untergruppe $\mathfrak{B}(U)$ von $\mathfrak{B}(H)$ homolog (weil die Elemente $S' \circ u$ von $\mathfrak{B}(H)$ in G homolog 0 sind).

Die Dimensionen der nicht zu der Untergruppe $\mathfrak{B}(U)$ von $\mathfrak{B}(G)$ gehörigen Elemente $v_{l'+1}, v_{l'+2}, \cdots, v_l$ müssen also $\geqq m$ sein; da die Summe der Dimensionen, wie oben festgestellt, gleich m ist, gibt es nur ein einziges solches Element, und es hat die Dimension m.

Das bedeutet: Man erhält eine Basis von $\mathfrak{B}(G)$, indem man zu einer Basis $\{v_1, v_2, \cdots, v_{l'}\}$ von $\mathfrak{B}(U)$ ein gewisses Element v_l mit $d(v_l) = m$ hinzufügt.

Im Falle eines ungeraden m ist nun das Element $\bar{v}$ ein minimales Element der geraden Dimension $m - 1$, und folglich gleich Null (Kap. I, §2, Nr. 3). Die eben angestellten Ueberlegungen sind also anwendbar auf diesen Fall, und enthalten offenbar einen Beweis der Behauptung a) aus Nr. 1.

Da aus $\bar{v} = 0$ die Existenz eines minimalen Elementes $v_l \neq 0$ mit $d(v_l) = m$ folgt, und andererseits ein minimales Element gerader Dimension immer Null ist, so muss im Fall eines geraden m das Element $\bar{v}$ von Null verschieden sein. Damit kommen wir zum Fall b).

b) Es sei $\bar{v} \neq 0$. Nimmt man $\bar{v}$ in eine Basis $\{v_1, v_2, \cdots, v_{l'-1}, \bar{v}\}$ von $\mathfrak{B}(U)$ auf, so sind, weil die rationalen Vielfachen von $\bar{v}$ die einzigen Elemente von $\mathfrak{B}(U)$ sind, die in G homolog 0 sind, die Elemente $v_1, v_2, \cdots, v_{l'-1}$ in G linear unabhängig und können durch Elemente $v_{l'}, v_{l'+1}, \cdots, v_l$ zu einer Basis von $\mathfrak{B}(G)$ ergänzt werden. Wegen $d(\bar{v}) = m - 1$ gilt

$$d(v_{l'}) + d(v_{l'+1}) + \cdots + d(v_l) = d(G) - (d(U) - d(v)) = 2m - 1.$$

Nun kann aber, genau wie in a), unter den Elementen $v_{l'}, v_{l'+1}, \cdots, v_l$ keines vorkommen, dessen Dimension kleiner als m ist; folglich kann es nur ein solches Element geben, und es muss die Dimension $2m - 1$ haben.

Das bedeutet: Es gibt in $\mathfrak{B}(U)$ ein Element $\bar{v}$ der Dimension $m - 1$, dass nicht Null ist; und aus einer Basis von $\mathfrak{B}(U)$, in der dieses Element $\bar{v}$ als Basiselement auftritt, erhält man eine Basis von $\mathfrak{B}(G)$, indem man $\bar{v}$ ersetzt durch ein gewisses Element v_l der Dimension $2m - 1$.

Aus $\bar{v} \sim 0$ in G folgt nach Kap. II, §2, Nr. 2, Korollar 1, dass auch $U \sim 0$ in G ist.

Damit ist die Behauptung b) von Nr. 1 vollständig bewiesen.

2. Die Gruppen A_n, B_n, C_n, D_n

Mit Hilfe des Satzes IV aus §1 lassen sich die Homologieringe der klassischen Gruppen A_n, B_n, C_n, D_n leicht bestimmen, auf Grund der Tatsache, dass

diesen Gruppen in natürlicher Weise Sphären als Wirkungsräume zugeordnet sind.

1. Die Gruppe A_n ($n = 1, 2, \cdots$) ist die Gruppe der linearen, unitären, unimodularen Transformationen in $n + 1$ Variabeln, also die Gruppe der Transformationen

$$z_j' = \sum a_{jk} z_k$$

mit der Determinante $|a_{jk}| = 1$, die die Hermite'sche Form

$$z_1 \cdot \bar{z}_1 + z_2 \cdot \bar{z}_2 + \cdots + z_{n+1} \cdot \bar{z}_{n+1}$$

invariant lassen [23]. Wir setzen $z_j = x_j + iy_j$ (x_j, y_j reell); die Hermite'sche Form geht dabei über in

$$x_1^2 + y_1^2 + x_2^2 + y_2^2 + \cdots + x_{n+1}^2 + y_{n+1}^2 .$$

Die Transformationen von A_n lassen sich also auffassen als (reelle) orthogonale Transformationen eines Euklidischen R_{2n+2} mit den Koordinaten x_1, y_1, x_2, y_2, $\cdots$, x_{n+1}, y_{n+1}, die paarweise zu den komplexen Koordinaten $z_j = x_j + iy_j$ zusammengefasst sind. A_n ist also eine Gruppe von topologischen (sogar orthogonalen) Transformationen der $(2n + 1)$-dimensionalen Sphäre S_{2n+1}, die durch

$$x_1^2 + y_1^2 + x_2^2 + y_2^2 + \cdots + x_{n+1}^2 + y_{n+1}^2 = 1$$

gegeben ist. Man sieht leicht, dass diese Gruppe die S_{2n+1} transitiv transformiert. Es genügt zu beweisen, dass man den Punkt ($z_1 = 1, z_i = 0$ ($i > 1$)) in einen beliebigen Punkt (a_1, a_2, $\cdots$, a_{n+1}) (mit $\sum a_j \cdot \bar{a}_j = 1$) überführen kann. Und das kann man, weil man immer unitäre, unimodulare Matrizen angeben kann, deren erste Spalte (a_{11}, a_{21}, $\cdots$, $a_{n+1,1}$) mit (a_1, a_2, $\cdots$, a_{n+1}) übereinstimmt. S_{2n+1} ist also ein Wirkungsraum von A_n (die Bedingung $T_p \cdot T_q = T_{p \cdot q}$ aus Kap. II, §2, Nr. 4 ist natürlich erfüllt). Die Isotropiegruppe ist die Untergruppe, die einen Punkt, etwa den Punkt ($z_1 = 1, z_i = 0$ ($i > 1$)) festhält; das ist offenbar die Gruppe A_{n-1}.

$2n + 1$ ist ungerade; man erhält also nach Satz IV a) (§1): Der Homologiering $\mathfrak{R}(A_n)$ ist isomorph dem Ring $\mathfrak{R}(A_{n-1} \times S_{2n+1})$. Die Mannigfaltigkeit A_1 ist bekanntlich der Sphäre S_3 homöomorph. Damit erhält man durch Induktion [24]:

Der Homologiering $\mathfrak{R}(A_n)$ ist isomorph dem Ring $\mathfrak{R}(\Pi)$ der Mannigfaltigkeit

$$\Pi = S_3 \times S_5 \times S_7 \times \cdots \times S_{2n+1} ;$$

dabei bezeichnet S_i die i-dimensionale Sphäre.

Definiert man für einen beliebigen Komplex k das Poincaré'sche Polynom $P_k(t)$ durch:

$$P_k(t) = p_0 + p_1 \cdot t + p_2 \cdot t^2 + \cdots ,$$

wobei p_i die i-te Betti'sche Zahl von k und t eine Variable ist, und beachtet, dass

$$P_{k_1 \times k_2}(t) = P_{k_1}(t) \cdot P_{k_2}(t)$$

ist, so liest man ab:

Das Poincaré'sche Polynom von A_n ist

$$P_{A_n}(t) = (1 + t^3) \cdot (1 + t^5) \cdot (1 + t^7) \cdots (1 + t^{2n+1}).$$

2. Die Gruppe B_n ist die orthogonale Gruppe in $2n + 1$ Variabeln [23]. Das ist eine transitive Transformationsgruppe der Sphäre S_{2n}; also ist S_{2n} ein Wirkungsraum von B_n.

Die Gruppe D_n ist die orthogonale Gruppe in $2n$ Variabeln [23]. Das ist eine transitive Transformationsgruppe der Sphäre S_{2n-1}; also ist S_{2n-1} ein Wirkungsraum von D_n.

B_n ist definiert für $n = 1, 2, \cdots$; D_n wird bei der Aufzählung der einfachen Gruppen nur für $n \geqq 3$ betrachtet; für unsere Zwecke können wir aber die Fälle $n = 1, 2$ in die Definition von D_n einbeziehen. D_1 ist dem Kreis S_1 homöomorph.

Die Isotropiegruppe des Wirkungsraumes S_{2n-1} von D_n ist die Gruppe B_{n-1}; das sieht man, wenn man als Punkt der Sphäre $\sum x_i^2 = 1$, den man festhält, den Punkt $(1, 0, 0, \cdots, 0)$ wählt. $2n - 1$ ist ungerade; aus Satz IV a) folgt also:

Der Ring $\mathfrak{R}(D_n)$ ist isomorph dem Ring $\mathfrak{R}(B_{n-1} \times S_{2n-1})$. Dabei ist $n > 1$ vorausgesetzt; $\mathfrak{R}(D_1)$ ist isomorph $\mathfrak{R}(S_1)$.

Die Isotropiegruppe des Wirkungsraumes S_{2n} von B_n ist die Gruppe D_n; das sieht man wieder, wenn man den Punkt $(1, 0, \cdots, 0)$ festhält. $2n$ ist gerade; also kann man Satz IV b) anwenden. Danach hat man $\mathfrak{R}(D_n)$ als $\mathfrak{R}(\Pi \times S_{2n-1})$ darzustellen, und erhält dann $\mathfrak{R}(B_n)$ als $\mathfrak{R}(\Pi \times S_{4n-1})$. Wegen des eben erhaltenen Resultats über $\mathfrak{R}(D_n)$ kann man B_{n-1} für Π einsetzen; damit ist gezeigt:

Der Ring $\mathfrak{R}(B_n)$ ist isomorph dem Ring $\mathfrak{R}(B_{n-1} \times S_{4n-1})$, für $n > 1$. Für $n = 1$, folgt, nach der Bemerkung nach Satz IV, aus der Isomorphie von $\mathfrak{R}(D_1)$ mit $\mathfrak{R}(S_1)$, dass $\mathfrak{R}(B_1)$ mit $\mathfrak{R}(S_3)$ isomorph ist; das ergibt sich auch daraus, dass B_1 bekanntlich mit dem projektiven Raum P_3 homöomorph ist.

Durch Induktion erhält man jetzt sofort [24]:

Der Homologiering $\mathfrak{R}(B_n)$ ist isomorph dem Ring $\mathfrak{R}(\Pi)$ der Mannigfaltigkeit

$$\Pi = S_3 \times S_7 \times S_{11} \times \cdots \times S_{4n-1};$$

Der Homologiering $\mathfrak{R}(D_n)$ ist isomorph dem Ring $\mathfrak{R}(\Pi)$ der Mannigfaltigkeit

$$\Pi = S_3 \times S_7 \times S_{11} \times \cdots \times S_{4n-5} \times S_{2n-1} \qquad (\textit{für } n > 1);$$

$\mathfrak{R}(D_1)$ ist isomorph $\mathfrak{R}(S_1)$.

Für die Poincaré'schen Polynome entnimmt man daraus:

$$P_{B_n}(t) = (1 + t^3)(1 + t^7) \cdots (1 + t^{4n-1});$$

$$P_{D_n}(t) = (1 + t^3)(1 + t^7) \cdots (1 + t^{4n-5})(1 + t^{2n-1}) \quad \textit{für } n > 1; P_{D_1}(t) = 1 + t.$$

3. Die Gruppe C_n ($n = 1, 2, \cdots$) ist die symplektische Gruppe [23, 25]: die Gruppe derjenigen unitären Transformationen eines $\bar{R}_{2n}$ mit den (komplexen) Koordinaten $(x_1, x_1', x_2, x_2', \cdots, x_n, x_n')$, die die schiefsymmetrische Bilinearform

$$(x_1y_1' - x_1'y_1) + (x_2y_2' - x_2'y_2) + \cdots + (x_ny_n' - x_n'y_n)$$

invariant lassen.

Den $\bar{R}_{2n}$ kann man, indem man die Koordinaten x_j, x_j' in Real- und Imaginärteil zerlegt, als einen reellen, Euklidischen, R_{4n} auffassen; die Transformationen von C_n sind dann wegen ihrer Unitarität in $\bar{R}_{2n}$ orthogonale Transformationen der $(4n - 1)$-dimensionalen Sphäre (etwa mit dem Radius 1) um den Nullpunkt des R_{4n}. Ist $(c_1, c_1', \cdots, c_n, c_n')$ ein beliebiger Punkt dieser Sphäre (ist also $c_1 \cdot \bar{c}_1 + c_1' \cdot \bar{c}_1' + \cdots + c_n' \cdot \bar{c}_n' = 1$), so kann man eine Matrix aus C_n angeben, deren erste Spalte $(c_{11}, c_{11}', c_{21}, c_{21}', \cdots, c_{n1}, c_{n1}')$ mit $(c_1, c_1', \cdots, c_n, c_n')$ übereinstimmt [25]; das bedeutet: die S_{4n-1} wird transitiv transformiert, sie ist ein Wirkungsraum von C_n.

Die zugehörige Isotropiegruppe ist C_{n-1}: Man halte den Punkt $(1, 0, 0, \cdots, 0)$ (im $\bar{R}_{2n}$) fest; dann rechnet man (unter Beachtung der Invarianz der Bilinearform und der Unitarität) sofort nach, dass auch der Punkt $(0, 1, 0, \cdots, 0)$ des $\bar{R}_{2n}$ festbleibt. Beachtet man noch einmal die Unitarität, so erkennt man: Die Isotropiegruppe besteht aus den Transformationen von C_n, für die

$$x_1 \rightarrow x_1, \qquad x_1' \rightarrow x_1'$$

gilt, und das ist eben die Gruppe C_{n-1}.

$4n - 1$ ist ungerade. Nach Satz IV a) ist also $\Re(C_n)$ isomorph mit $\Re(C_{n-1} \times S_{4n-1})$. Beachtet man noch, dass die Gruppe C_1 mit der Gruppe A_1 identisch ist, so erhält man durch Induktion [24]:

Der Homologiering $\Re(C_n)$ ist isomorph dem Ring $\Re(\Pi)$ der Mannigfaltigkeit

$$\Pi = S_3 \times S_7 \times S_{11} \times \cdots \times S_{4n-1}$$

Daraus liest man ab: *Das Poincaré'sche Polynom von C_n ist*

$$P_{C_n}(t) = (1 + t^3)(1 + t^7) \cdots (1 + t^{4n-1}).$$

Kapitel IV

Homologieeigenschaften der Wirkungsräume

Nachdem wir im Kapitel III den speziellen Fall $W = S_m$ behandelt haben, wollen wir jetzt zur Untersuchung beliebiger Wirkungsräume übergehen. Die Untersuchung, die wir im Folgenden durchführen, ermöglicht es, wenn $U \nsim 0$ in G ist, den Ring $\Re(W)$ durch $\Re(G)$ und $\Re(U)$ zu bestimmen.

1. Die Gruppe $P(\mathfrak{B}(G))$

1. Wie immer, bedeute G eine Gruppe, U eine (zusammenhängende) Untergruppe, W den zugehörigen Wirkungsraum G/U. Dann ist W eine orientierbare, geschlossene Mannigfaltigkeit der Dimension $d(W) = d(G) - d(U)$. Die Projektion P von G auf W bewirkt eine Abbildung P der Betti'schen Gruppe $\mathfrak{B}(G)$ in die Betti'sche Gruppe $\mathfrak{B}(W)$. Die Gruppe derjenigen Elemente von $\mathfrak{B}(W)$, die vermöge P Bilder von Elementen von $\mathfrak{B}(G)$ sind, heisse die Bildgruppe $P(\mathfrak{B}(G))$; ihre Struktur soll untersucht werden.

Wenn U ein Normalteiler von G ist, dann ist P die natürliche homomorphe Abbildung von G auf die Faktorgruppe G/U, hat also die in Kap. II, §1 beschriebenen Eigenschaften. Wir werden nun zeigen, dass auch dann, wenn U kein Normalteiler ist, die Abbildung P von diesem Typ ist; nur muss berücksichtigt werden, dass es in W im allgemeinen keine Multiplikation gibt.

2. Es sei also U eine beliebige (zusammenhängende) Untergruppe von G. Wir beweisen dann den

Satz V. *Der Kern der Abbildung P von $\mathfrak{B}(G)$ in $\mathfrak{B}(W)$ ist das Ideal in $\mathfrak{P}(G)$ (aufgefasst als Untergruppe von $\mathfrak{B}(G)$), das erzeugt wird von denjenigen minimalen Elementen von G, deren Bild bei P die Null ist.*

Die Restklassengruppe von $\mathfrak{B}(G)$ nach diesem Ideal (das Ideal als Untergruppe von $\mathfrak{B}(G)$ aufgefasst) wird also isomorph auf die Bildgruppe abgebildet. (Die Elemente dieser Restklassengruppe bilden zugleich den Restklassenring von $\mathfrak{P}(G)$ nach dem Ideal.)

Der Beweis geht aus von dem Aufspannsatz in der Form I (Kap. I, §3, Nr. 2) und wird folgendermassen geführt:

Die Gruppe $\mathfrak{B}(G)$ ist durch P homomorph in $\mathfrak{B}(W)$ abgebildet. $\mathfrak{B}^0(G)$ sei der Kern dieser Abbildung; $\mathfrak{B}(G)$ werde als direkte Summe $\mathfrak{B}^0(G) + \mathfrak{B}^1(G)$ dargestellt; $\mathfrak{B}^1(G)$ wird also isomorph abgebildet. $\{v_1, v_2, \cdots, v_m\}$ sei eine Basis von $\mathfrak{B}^1(G)$, und $\{v_{m+1}, v_{m+2}, \cdots, v_l\}$ eine von $\mathfrak{B}^0(G)$; dann ist $\{v_1, v_2, \cdots, v_l\}$ eine Basis von $\mathfrak{B}(G)$. Die Produkte

$$v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} \qquad (i_1 < i_2 < \cdots < i_r)$$

bilden mit v_0 zusammen eine Basis $\langle v_i \rangle$ von $\mathfrak{B}(G)$ nach Satz I. Dann beweisen wir:

a) Ein solches Produkt wird durch P auf 0 gebildet, wenn wenigstens ein Faktor v_{i_k} zu $\mathfrak{B}^0(G)$ gehört, wenn also wenigstens ein Index $i_k > m$ ist;

b) diejenigen Produkte, in denen alle Faktoren v_{i_k} zu $\mathfrak{B}^1(G)$ gehören, in denen also alle Indizes $i_k \leqq m$ sind, werden durch P auf linear unabhängige Elemente von $\mathfrak{B}(W)$ abgebildet. Aus a) und b) folgt dann sofort der Satz V: Die in a) auftretenden Elemente bilden eine Basis des von $\mathfrak{B}^0(G)$ erzeugten Ideals von $\mathfrak{P}(G)$; dieses Ideal ist das in Satz V genannte. Aus a) folgt, dass das von $\mathfrak{B}^0(G)$ erzeugte Ideal durch P auf 0 abgebildet wird; aus b) entnimmt man, dass nur die Elemente dieses Ideals auf 0 abgebildet werden, womit Satz V bewiesen ist.

Man kann b) auch so formulieren: der von v_0 und $v_1, v_2, \cdots, v_m$ erzeugte Teilring von $\mathfrak{P}(G)$ wird eineindeutig abgebildet; wegen a) wird er *auf* die Bildgruppe abgebildet. Bezeichnet S_i eine Sphäre mit $d(S_i) = d(v_i)$ für $i = 1, 2, \cdots, m$, so sieht man daraus: Die Gruppe $P(\mathfrak{B}(G))$ ist isomorph der Betti'schen Gruppe der Mannigfaltigkeit $S_1 \times S_2 \times S_3 \times \cdots \times S_m$.

Aus a) und b) kann man noch folgenden Schluss ziehen, der den Charakter der Abbildung P besser verstehen lässt: $v_1', v_2', \cdots, v_r'$ seien irgendwelche Elemente von $\mathfrak{B}(G)$; das Bild $P(v_1' \circ v_2' \circ \cdots \circ v_r')$ von $v_1' \circ v_2' \circ \cdots \circ v_r'$ ist dann und nur dann ~ 0 in W, wenn die Elemente $\eta_i' = P(v_i')$ in W linear abhängig sind.

Sind nämlich die η_i' unabhängig, so kann man die v_i' in eine Basis von $\mathfrak{B}^1(G)$ aufnehmen, und nach b) ist $P(v_1' \circ v_2' \circ \cdots \circ v_r')$ nicht Null.

Sind aber die η_i' abhängig, so kann man durch eine lineare Transformation der v_i' erreichen, dass $\eta_r' = P(v_r')$ gleich Null ist; das Produkt $v_1' \circ v_2' \circ \cdots \circ v_r'$ ist nach Kap. I, §3, Nr. 9 invariant gegen eine solche Transformation, und aus $P(v_r') = 0$ folgt nach a), dass auch $P(v_1' \circ v_2' \circ \cdots \circ v_r') = 0$ ist.

Wir kommen jetzt zum Beweis unseres Satzes.

3. Zuerst beweisen wir den Teil a).

Wir definieren eine Abbildung T des topologischen Produktes $G \times W$ auf W, indem wir setzen:

$$T(p \times q') = T_p(q'),$$

wo p ein Punkt von G, und q' ein Punkt von W ist; T_p ist erklärt in Kap. II, §2, Nr. 4. Schreibt man W als Nebengruppenraum, und bedeutet q' die Nebengruppe qU, dann ist $T(p \times q')$ die Nebengruppe pqU.

Der Punkt $T(p \times q')$ heisst das Produkt von p und q' und wird auch mit pq' bezeichnet. (Wenn Produkte $c \cdot d$ auftreten, so wird es immer klar sein, ob es sich um das Gruppenprodukt in G handelt, wenn nämlich $c, d \in G$, oder um das eben definierte Produkt, wenn nämlich $c \in G$, $d \in W$; das gleiche gilt für das sofort zu definierende verallgemeinerte Pontrjagin'sche Produkt; es werden deshalb die alten Zeichen für die Multiplikation benützt.)

Ueber dieses Produkt sei bemerkt: $p \cdot q'$ ist bei festem p als Abbildung von W in sich aufzufassen, nämlich als $T_p(q')$; diese Abbildung ist topologisch und zur Identität homotop—man lasse p nach e, dem Einheitspunkt von G, wandern.

Bei festem q' dagegen ist $p \cdot q'$ eine Abbildung von G auf W, die zu der Projektion P homotop ist: man lasse q' in den die Nebengruppe U selbst darstellenden Punkt e' von W wandern; $p \cdot e'$ ist dann die Nebengruppe pU, als Punkt von W betrachtet, und das ist eben $P(p)$.

Die Eigenschaft des Produktes $p \cdot q'$, die wir brauchen, ist eine Funktionalgleichung für P; es gilt nämlich:

$$P(p_1 \cdot p_2) = p_1 \cdot P(p_2);$$

p_1, p_2 sind zwei beliebige Punkte von G; links steht die Projektion des Produktes $p_1 \cdot p_2$; rechts steht das Produkt des Punktes p_1 von G mit dem Punkte

$P(p_2)$ von W, also $T(p_1 \times P(p_2))$; beide Seiten der Gleichung bedeuten die Nebengruppe $p_1 \cdot p_2 \cdot U$, als Punkt von W.

Die Abbildung T bewirkt eine Abbildung T der Homologieklassen. Wir definieren nun, wie ja nahe liegt, das "verallgemeinerte Pontrjagin'sche Produkt $x \circ y'$" eines Elementes x von $\mathfrak{B}(G)$ mit einem Element y' von $\mathfrak{B}(W)$, indem wir setzen:

$$x \circ y' = T(x \times y');$$

das ist wieder ein Element von $\mathfrak{B}(W)$.

Diese Multiplikation ist offenbar distributiv mit der Addition verknüpft. Es gilt

$$x \circ 0' = 0 \circ y' = 0',$$

wo 0, bzw. $0'$, das Nullelement von $\mathfrak{B}(G)$, bzw. $\mathfrak{B}(W)$, ist, und x, y' beliebig sind.

Die Funktionalgleichung für P gibt für die Homologieklassen die entsprechende Funktionalgleichung, nämlich:

$$P(x_1 \circ x_2) = x_1 \circ P(x_2),$$

wenn x_1, x_2 zwei beliebige Elemente von $\mathfrak{B}(G)$ sind. Das ergibt sich so: F sei die bekannte Abbildung von $G \times G$ in G; mit S bezeichne man die Abbildung von $G \times G$ in $G \times W$, die durch $S(p_1 \times p_2) = p_1 \times P(p_2)$ gegeben ist. Dann ist wegen der Funktionalgleichung die Abbildung PF von $G \times G$ in W gleich der Abbildung TS. Nun ist $PF(x_1 \times x_2) = P(x_1 \circ x_2)$; andererseits ist $S(x_1 \times x_2) = x_1 \times P(x_2)$ (vgl. Kap. I, §1, Nr. 8), also $TS(x_1 \times x_2) = x_1 \circ P(x_2)$.

Jetzt folgt sofort die Behauptung a). $v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r}$ sei ein Element der Basis $\langle v_i \rangle$, und es sei $P(v_{i_k}) = 0'$, $k \leqq r$. Wegen der Antikommutativität der v_i darf man $k = r$ annehmen, und kann also das Element in der Form $w \circ v_{i_r}$ mit $w = v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_{r-1}}$ schreiben. Dann ist

$$P(w \circ v_{i_r}) = w \circ P(v_{i_r}) = w \circ 0' = 0'.$$

4. Jetzt beweisen wir die Behauptung b); wir benützen den Umkehrungshomomorphismus φ der Abbildung P.

Als Basis in $\mathfrak{B}(G)$ benützen wir die von v_0 und den Elementen

$$v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} \qquad (i_1 < i_2 < \cdots < i_r)$$

gebildete Basis $\langle v_i \rangle$. Nach dem Aufspannsatz (vgl. Kap. I, §3, Nr. 9) ist die duale Basis die Basis $[z_i]$, die erzeugt wird von den zu den v_i dualen Elementen z_i.

Die Bilder $\eta_i = P(v_i)$ der Elemente $v_1, v_2, \cdots, v_m$ sind nach Voraussetzung in W linear unabhängig. Sie werden in eine Betti'sche Basis von $\mathfrak{B}(W)$ aufgenommen; ζ_i $(i = 1, 2, \cdots, m)$ seien die in der dualen Basis zu den η_i dualen Elemente.

Wir betrachten $\varphi(\zeta_i)$. Wir behaupten: stellt man $\varphi(\zeta_i)$ in der Basis $[z_i]$ dar, dann ist z_i das einzige maximale Basiselement, das in $\varphi(\zeta_i)$ auftritt, und zwar hat es den Koeffizienten 1. Wir behaupten also:

$$(\zeta) \qquad \varphi(\zeta_i) = z_i + \sum a_h x_h,$$

wo die x_h zusammengesetzte Elemente der Basis $[z_i]$ und die a_h Koeffizienten sind.

Zum Beweis betrachten wir das Produkt $\varphi(\zeta_i)\cdot v_k$. Die Funktionalgleichung (Kap. I, §1, Nr. 5) für φ liefert

$$P(\varphi(\zeta_i)\cdot v_k) = \zeta_i \cdot P(v_k) = \zeta_i \cdot \eta_k, \qquad \text{bzw.} = \zeta_i \cdot 0'$$

also

$$P(\varphi(\zeta_i)\cdot v_k) = \delta_{ik}\cdot \eta_0 \qquad (i = 1, \cdots, m;\, k = 1, \cdots, l),$$

wo $\eta_0 = P(v_0)$ der einfach gezählte Punkt von W ist; denn weil η_k minimal ist, ist $\zeta_k \cdot \eta_k = \eta_0$ der einzige von Null verschiedene Schnitt von η_k, für $k = 1, 2, \cdots, m$; für $k > m$ ist $P(v_k) = 0'$. Also ist

$$\varphi(\zeta_i)\cdot v_k = \delta_{ik} v_0 ;$$

daraus folgt auf Grund der Rechenregeln in $\mathfrak{R}(G)$ sofort die Behauptung (ζ)—man beachte etwa die Gleichungen $z_i \cdot v_k = \delta_{ik} v_0$ $(i, k = 1, \cdots, l)$ (Kap. I, §3, Nr. 8).

Wegen der Multiplikativität von φ hat man dann:

$$\varphi(\zeta_{i_1}\cdot\zeta_{i_2}\cdot \cdots \cdot\zeta_{i_r}) = z_{i_1}\cdot z_{i_2}\cdot \cdots \cdot z_{i_r} + \sum y_h ,$$

wo jedes y_h ein Produkt von mindestens $r + 1$ Faktoren z_i ist.

Aus dieser Darstellung leitet man die folgenden Formeln ab (vgl. Kap. I, §3, Nr. 9):

1) $$\varphi(\zeta_{i_1}\cdot\zeta_{i_2}\cdot \cdots \cdot\zeta_{i_r})\cdot(v_{i_1}\circ v_{i_2}\circ \cdots \circ v_{i_r}) = \pm v_0 ,$$

2) $$\varphi(\zeta_{i_1}\cdot\zeta_{i_2}\cdot \cdots \cdot\zeta_{i_r})\cdot(v_{j_1}\circ v_{j_2}\circ \cdots \circ v_{j_r}) = 0,$$

wenn die Indexmengen $\{i_1, i, \cdots, i_r\}$ und $\{j_1, j_2, \cdots, j_r\}$ voneinander verschieden sind,

3) $$\varphi(\zeta_{i_1}\cdot\zeta_{i_2}\cdot \cdots \cdot\zeta_{i_r})\cdot(v_{k_1}\circ v_{k_2}\circ \cdots \circ v_{k_s}) = 0, \qquad \text{wenn } s < r.$$

5. Daraus folgt nun leicht die Behauptung b) in der Formulierung: der von $v_0, v_1, v_2, \cdots, v_m$ erzeugte Teilring von $\mathfrak{P}(G)$ wird eineindeutig abgebildet.

Es sei nämlich w ein Element dieses Teilringes, also eine Linearkombination $\sum a_i w_i$ (mit Koeffizienten a_i), in der die w_i solche Elemente[6]

$$v_{i_1}\circ v_{i_2}\circ \cdots \circ v_{i_r} \qquad (i_1 < i_2 < \cdots < i_r),$$

bedeuten, in denen $i_k \leqq m$ gilt; die Koeffizienten a_i seien von Null verschieden. $w_1 = v_{h_1}\circ v_{h_2}\circ \cdots \circ v_{h_t}$ sei eines dieser Elemente mit maximaler Anzahl von Faktoren v_i. Dann ist nach Nr. 4, 1),

$$\varphi(\zeta_{h_1}\cdot\zeta_{h_2}\cdot \cdots \cdot\zeta_{h_t})\cdot w_1 = \pm v_0 ,$$

und nach Nr. 4, 2) und 3),

$$\varphi(\zeta_{h_1}\cdot\zeta_{h_2}\cdot \cdots \cdot\zeta_{h_t})\cdot w_i = 0 \qquad \text{für } i > 1:$$

also ist schliesslich

$$\varphi(\zeta_{h_1}\cdot\zeta_{h_2}\cdot \cdots \cdot\zeta_{h_t})\cdot w = \pm a_1 v_0 .$$

[6] Wir dürfen $w_i \neq v_0$ annehmen.

Nach der Funktionalgleichung für φ folgt hieraus

(1) $$(\zeta_{h_1}\cdot\zeta_{h_2}\cdot\ \cdots\ \cdot\zeta_{h_t})\cdot P(w) = \pm a_1\eta_0 .$$

Also ist $P(w) \neq 0$, q.e.d.

Es sei noch bemerkt, dass das Produkt aller ζ_i, das Element $\zeta_1\cdot\zeta_2\cdot\ \cdots\ \cdot\zeta_m$, von Null verschieden ist: man setze für w das Produkt $v_1\circ v_2\circ\ \cdots\ \circ v_m$ ein.

2. $U \nsim 0$

1. Für den Fall, dass $U \nsim 0$ in G ist, lässt sich jetzt eine sehr präzise Aussage herleiten. Wir beweisen nämlich den folgenden Satz:

Satz VI. *Die Untergruppe U der Gruppe G sei nicht homolog* 0 *in G; der zugehörige Wirkungsraum sei $G/U = W$; dann gilt*:

a) *der Ring $\mathfrak{R}(W)$ ist isomorph dem Ring eines topologischen Produktes von Sphären ungerader Dimensionen,*

b) *der Ring $\mathfrak{R}(G)$ ist isomorph dem Ring $\mathfrak{R}(U \times W)$ des topologischen Produktes $U \times W$.*

Danach ist $\mathfrak{R}(W)$ durch $\mathfrak{R}(G)$ und $\mathfrak{R}(U)$ bestimmt: $l_i(W)$, die Zahl der i-dimensionalen Sphären in dem $\mathfrak{R}(W)$ darstellenden Sphärenprodukt, ist gegeben durch $l_i(G) - l_i(U)$.

Zum Beweis zeigen wir erstens, dass die Projektion P eine Abbildung von $\mathfrak{R}(G)$ *auf* $\mathfrak{R}(W)$ liefert; nach §1 ist dann $\mathfrak{R}(W)$ eineindeutiges Bild eines gewissen Teilringes des Pontrjagin'schen Ringes $\mathfrak{P}(G)$. Zweitens bestimmen wir diesen Teilring; er ergibt sich als isomorph dem Restklassenring des von den Elementen von $\mathfrak{B}(U)$ erzeugten Ideals von $\mathfrak{P}(G)$. Daraus folgt dann leicht der Satz.

2. U ist $\nsim 0$ in G; die Homologieklasse von G, in der U liegt, bezeichnen wir auch mit U. Unter φ verstehen wir, wir in §1, den Umkehrungshomomorphismus der Projektion P. Dann gilt:

$$\varphi(\eta_0) = U.$$

Dabei ist η_0 der einfach gezählte Punkt in W.

Es gilt nämlich allgemein für die Projektion eines gefaserten Raumes auf den Faserraum: Der Urbildzyklus (das φ-Bild) des Punktes ist die Faser. Das folgt z.B. daraus, dass man die Umkehrung φ einer solchen Abbildung P folgendermassen erklären kann: Der Faserraum B sei hinreichend fein trianguliert. Für ein Simplex x_i von B definiert man $\varphi(x_i)$ als die, in geeigneter Weise als Komplex aufgefasste, Menge $P^{-1}(x_i)$, die fasertreu homöomorph mit $x_i \times F$ (F = Faser) ist. Für Komplexe, Zyklen usw. erhält man φ dann durch Addition.

Wegen der Voraussetzung $U \nsim 0$ in G ist also $\varphi(\eta_0) \neq 0$. Nach dem Hilfssatz aus Kap. I, §1, Nr. 6 ist dann P eine Abbildung von $\mathfrak{R}(G)$ *auf* $\mathfrak{R}(W)$.

Zieht man jetzt den Satz V heran, so sieht man, dass die Gruppe $\mathfrak{B}(W)$ bestimmt ist, wenn man die Gruppe $\mathfrak{B}^0(G)$, d.h. die Gruppe derjenigen mini-

malen Elemente von $\mathfrak{R}(G)$, deren P-Bild in W homolog 0 ist kennt. Diese Gruppe soll jetzt bestimmt werden.

3. Weil U $\nsim$ 0 ist, ist kein Element (ausser 0) von $\mathfrak{B}(U)$ homolog 0 in G (Kap. II, §2, Nr. 2, Korollar 1), und man kann $\mathfrak{B}(U)$ als Untergruppe von $\mathfrak{B}(G)$ auffassen. Die Elemente von $\mathfrak{B}(U)$ werden bei P sicher auf 0 abgebildet: die ganze Untergruppe U wird ja auf einen Punkt abgebildet. Wir behaupten: mit den Elementen von $\mathfrak{B}(U)$ sind die Elemente von $\mathfrak{B}(G)$, deren P-Bild die Null ist, erschöpft, oder in der Bezeichnung von §1, Nr. 2: es ist $\mathfrak{B}^0(G) = \mathfrak{B}(U)$.

Zum Beweis setzen wir $l(G) = l$ und $l(U) = l'$. Die Elemente $v_1, v_2, \cdots, v_{l'}$ mögen eine Basis von $\mathfrak{B}(U)$ bilden; sie werden aufgenommen in eine Basis $\{v_1, v_2, \cdots, v_l\}$ von $\mathfrak{B}(G)$. Dann gilt (nach dem Aufspannsatz, angewendet auf die Gruppe U; vgl. Kap. I, §3, Nr. 7)

$$v_1 \circ v_2 \circ \cdots \circ v_{l'} = U$$

als Homologie in U (dann bedeutet U die Eins von $\mathfrak{R}(U)$); also gilt die Gleichung auch als Homologie in G. Wir setzen noch

$$v_{l'+1} \circ v_{l'+2} \circ \cdots \circ v_l = V.$$

Dann gilt, wieder nach den Aufspannsatz, die Gleichung

$$U \circ V = 1,$$

wo 1 die Eins von $\mathfrak{R}(G)$ bedeutet; daher gilt (nach Kap. I, §3, Nr. 10) für den Schnitt $U \cdot V$:

$$U \cdot V = v_0 .$$

Uebt man auf diese Gleichung die Abbildung P aus, und beachtet, dass $U = \varphi(\eta_0)$ ist, so erhält man wegen der Funktionalgleichung für φ die Gleichung:

$$\eta_0 \cdot P(V) = \eta_0 .$$

Bezeichnet man die Eins von $\mathfrak{R}(W)$ mit $\bar{1}$, so heisst das:

$$P(V) = \bar{1}.$$

Es ist also $P(v_{l'+1} \circ v_{l'+2} \circ \cdots \circ v_l) \not\sim 0$; daraus folgt nach der letzten Bemerkung in §1, Nr. 2, dass die Bilder $\eta_i = P(v_i)$ der Elemente $v_{l'+1}, v_{l'+2}, \cdots, v_l$ in W linear unabhängig sind. Das bedeutet aber: die Restklassengruppe $\mathfrak{B}(G) - \mathfrak{B}(U)$ wird durch P isomorph abgebildet.

Daher ist $\mathfrak{B}(U)$ der Kern der Abbildung P von $\mathfrak{B}(G)$; die Behauptung $\mathfrak{B}^0(G) = \mathfrak{B}(U)$ ist bewiesen.

Die von $v_{l'+1}, v_{l'+2}, \cdots, v_l$ erzeugte Untergruppe von $\mathfrak{B}(G)$ ist die in §1, Nr. 2 mit $\mathfrak{B}^1(G)$ bezeichnete Gruppe; der von dem Punkt v_0 und den Elementen dieser Gruppe erzeugte Teilring von $\mathfrak{P}(G)$ wird nach §1, Nr. 2 eineindeutig auf die Bildgruppe abgebildet. Da jetzt nach Nr. 2 die Bildgruppe mit der ganzen

Gruppe $\mathfrak{B}(W)$ zusammenfällt, ist also $\mathfrak{B}(W)$ eineindeutiges Bild des von den $v_{l'+1}, v_{l'+2}, \cdots, v_l$ und v_0 erzeugten Teilringes von $\mathfrak{P}(G)$.

Dieser Teilring enthält nun genau $2^{l-l'}$ linear unabhängige Elemente, z.B. den Punkt v_0 und die Produkte

$$v_{i_1} \circ v_{i_2} \circ \cdots \circ v_{i_r} \qquad (l' + 1 \leqq i_k \leqq l, i_1 < i_2 < \cdots < i_r).$$

Also enthält auch $\mathfrak{B}(W)$ genau $2^{l-l'}$ linear unabhängige Elemente.

4. Wir betrachten die zu den $\eta_i = P(v_i)$ dualen Elemente ζ_i von $\mathfrak{B}(W)$ $(i = l' + 1, \cdots, l)$. Weil $P(V) = P(v_{l'+1} \circ v_{l'+2} \circ \cdots \circ v_l) = \bar{1}$ die Dimension $d(W)$ hat, ist der Schnitt $\zeta_{l'+1} \cdot \zeta_{l'+2} \cdot \cdots \cdot \zeta_l$ nulldimensional. Aus der Formel (1) in §1, Nr. 5, entnimmt man leicht, dass dieser Schnitt gleich $\pm \eta_0$, also von Null verschieden ist; man setze dazu für das dort mit w bezeichnete Element das Produkt $V = v_{l'+1} \circ v_{l'+2} \circ \cdots \circ v_l$ ein und beachte, dass $P(V) = \bar{1}$ ist. Ausserdem sind die ζ_i antikommutativ in Bezug auf den Schnitt:

$$\zeta_i \cdot \zeta_k = -\zeta_k \cdot \zeta_i \quad \text{und} \quad \zeta_i \cdot \zeta_i = 0,$$

denn es ist $\delta(\zeta_i) = d(\eta_i) = d(v_i)$ ungerade.

Daraus folgert man (vgl. die Betrachtung in Kap. I, §3, Nr. 7), dass die Produkte

$$\zeta_{i_1} \cdot \zeta_{i_2} \cdot \cdots \cdot \zeta_{i_r} \qquad (l' + 1 \leqq i_k \leqq l, i_1 < i_2 < \cdots < i_r)$$

linear unabhängige Elemente in $\mathfrak{B}(W)$ sind; mit $\bar{1}$ zusammen sind das genau $2^{l-l'}$ Elemente; die ζ_i bilden also mit $\bar{1}$ zusammen ein irreduzibles Erzeugendensystem von $\mathfrak{R}(W)$. Diese Eigenschaften der ζ_i sind es aber gerade, die $\mathfrak{R}(W)$ als Ring eines Produktes von ungerade-dimensionalen Sphären charakterisieren. Damit ist VI a) bewiesen. Den "Sphären selbst" entsprechen die Elemente η_i.

5. Den Isomorphismus von $\mathfrak{R}(G)$ und $\mathfrak{R}(U \times W)$ stellt man jetzt folgendermassen her:

1 sei die Eins von $\mathfrak{R}(G)$, $1'$ die von $\mathfrak{R}(U)$, $\bar{1}$ die von $\mathfrak{R}(W)$.

z_i seien die zu den v_i dualen Elemente von $\mathfrak{B}(G)$ $(i = 1, \cdots, l)$; sie bilden mit 1 zusammen ein irreduzibles Erzeugendensystem von $\mathfrak{R}(G)$ (vgl. Kap. I, §2, Nr. 4).

z'_j seien die zu den v_j dualen Elemente von $\mathfrak{B}(U)$ $(j = 1, \cdots, l')$; sie bilden mit $1'$ zusammen ein irreduzibles Erzeugendensystem von $\mathfrak{R}(U)$.

ζ_k seien die oben eingeführten Elemente von $\mathfrak{B}(W)$ $(k = l' + 1, \cdots, l)$; sie bilden mit $\bar{1}$ zusammen ein irreduzibles Erzeugendensystem von $\mathfrak{R}(W)$.

Die Elemente $(z'_j \times \bar{1})$ und $(1' \times \zeta_k)$ bilden dann mit dem Element $(1' \times \bar{1})$ zusammen ein irreduzibles Erzeugendensystem von $\mathfrak{R}(U \times W)$.

Alle diese Systeme sind antikommutativ; und das Produkt aller Elemente jeden Systems ist der einfach gezählte Punkt des betreffenden Ringes.

Jetzt ordnet man zu:

Den Elementen z_i $(i = 1, \cdots, l')$ die Elemente $(z'_i \times \bar{1})$,

den Elementen z_i ($i = l' + 1, \cdots, l$) die Elemente $(1' \times \zeta_k)$, dem Element 1 das Element $1' \times \bar{1}$.

Das ergibt offenbar einen Isomorphismus zwischen $\mathfrak{R}(G)$ und $\mathfrak{R}(U \times W)$ der gewünschten Art; damit ist auch VI b) bewiesen.

ZÜRICH, SWITZERLAND

LITERATURVERZEICHNIS

1. L. Pontrjagin: *Topological Groups* (Princeton 1939), Def. 38, S. 181 u. S. 83, D).
2. E. Cartan: *La Topologie des . . . Groupes de Lie* [L'Enseignement Math. 35 (1936); dasselbe auch: Actualités Scient. et Industr. 358].
3. L. Pontrjagin: *On Homologies in compact Lie groups* [Recueil math. de Moscou, t. 6 (48), 3 (1939)].
4. R. Brauer: C. R. 201 (1935), S. 419.
5. Ch. Ehresmann: C. R. 208 (1939), S. 321 u. S. 1263.
6. A. a. O. [2], S. 199.
7. H. Hopf: *Ueber die Topologie der Gruppen-Mannigfaltigkeiten und ihrer Verallgemeinerungen* [these Annals (2) 42 (1941)], Satz I.
8. A. a. O. [2], S. 198. Diese Vermutung ist auch ausgesprochen bei Hopf, a. a. O. [7], Nr. 37.
9. E. Cartan: *La theorie des groupes finis et continus et l'analysis situs* [Memorial Sc. Math. XLII], Nr. 17.
10. J. Feldbau: C. R. 208 (1939), S. 1621.
11. S. Lefschetz: *Topology* (New York 1930).
12. H. Hopf: *Zur Algebra der Abbildungen von Mannigfaltigkeiten* [Journal f. d. r. u. a. Math. 163 (1930)].
    H. Freudenthal: *Zum Hopf'schen Umkehrhomomorphismus* [Annals of Math. (2) 38 (1937)].
    A. Komatu: *Über die Ringdualität eines Kompaktums* [Tôhoku Math. J. 43 (1937)].
    H. Whitney: *On Products in a complex* [Annals of Math. (2) 39 (1938)], Th. 6.
13. Alexandroff-Hopf: *Topologie I* (Berlin 1935), Kap. VII, §3.
14. A. a. O. [11].
15. A. a. O. [2], S. 198.
16. A. a. O. [7], Nr. 37.
17. A. a. O. [1], Th. 50, insbesondere S. 200, Formel (20).
18. S. S. Cairns: *On the triangulation of regular loci* [Annals of Math. (2) 35 (1934)].
    H. Freudenthal: *Die Triangulation der differenzierbaren Mannigfaltigkeiten* [Proceedings Amsterdam XLII (1939)].
19. A. a. O. [10], Th. A.
20. A. a. O. [13], S. 128 u. 149.
21. A. a. O. [13], S. 334.
22. A. a. O. [13], Kap. VII, §2.
23. A. a. O. [3], S. 397 f.
24. A. a. O. [3], Th. 2 u. 3.
25. H. Weyl: *The classical groups* (Princeton 1940), S. 163 u. S. 172.

ANNALS OF MATHEMATICS
Vol. 42, No. 5, December, 1941

INTRINSIC TOPOLOGY AND COMPLETION OF BOOLEAN RINGS

BY HENRY LÖWIG

(Received October 29, 1940)

NOTATION

If $\mathfrak{B}$ is a subclass of a lattice A we denote the meet (cross-cut, intersection, product) of the elements of $\mathfrak{B}$ with respect to A, i. e. the maximal element of A, contained in all elements of $\mathfrak{B}$, its existence supposed, by $\prod_{x \in \mathfrak{B}}^{(A)} x$, and the join (union, conjunction, lattice-sum) of the elements of $\mathfrak{B}$ with respect to A, i.e. the minimal element of A, containing all elements of $\mathfrak{B}$, its existence supposed, by $\sum_{x \in \mathfrak{B}}^{(A)} x$. If $\mathfrak{B}$ consists only of two elements, a and b, we write also $a \wedge_A b$ instead of $\prod_{x \in \mathfrak{B}}^{(A)} x$ and $a \vee_A b$ instead of $\sum_{x \in \mathfrak{B}}^{(A)} x$. If A is especially a Boolean ring we denote the ring-sum (the symmetric difference) of two elements a and b with respect to A by $a \triangle_A b$. We omit the superscripts of the signs $\prod^{(A)}$ and $\sum^{(A)}$, and the subscript of the sign $\vee_A$ if there is no doubt to which lattice or Boolean ring they refer. Under the same condition we write $a \cdot b$ or ab instead of $a \wedge_A b$, and $a + b$ instead of $a \triangle_A b$. We call a subclass $\mathfrak{B}$ of a lattice A *bounded* if there exists at least one couple a, b of elements of A such that

$$a < x < b$$

for all elements x of $\mathfrak{B}$. The expressions of "σ-lattice" and "complete lattice" are used in the same sense as in the paper (VIII). (See especially (VIII), p. 795, 4th passage.) We call a Boolean ring complete if it is a complete lattice, and a Boolean σ-ring if it is a σ-lattice. If a subring (an ideal) of a Boolean ring is a σ-lattice we call it a σ-subring (a σ-ideal). The expression of "dual ideal" is used in the same sense as in (VI). The words "A Boolean ring A is given, and the small Latin letters ... denote elements of A" will be frequently suppressed. If we consider a sequence of elements a_n of a lattice we shall mostly suppress the supposition that the index n runs through all natural numbers. Otherwise we shall use the terminology and denotation of M. H. Stone's paper (V). The results of this paper will be used without any further reference.

1. Introduction

A sequential topology of a lattice can be defined in the following way:

If a sequence of elements a_n of a lattice has the property that all products $\prod_{k=n}^{\infty} a_k$ $(n = 1, 2, 3, \cdots)$, all joins $\sum_{k=n}^{\infty} a_k$ $(n = 1, 2, 3, \cdots)$, the join $\sum_{n=1}^{\infty} \prod_{k=n}^{\infty} a_k$, and the product $\prod_{n=1}^{\infty} \sum_{k=n}^{\infty} a_k$ exist, and

$$\sum_{n=1}^{\infty} \prod_{k=n}^{\infty} a_k = \prod_{n=1}^{\infty} \sum_{k=n}^{\infty} a_k = a,$$

the sequence a_n is said to converge towards a or to have the limit a. (See for instance (II), p. 453.)

The lattices considered in L. V. Kantorovitch's paper (VII) have even the property that $\prod_{k=n}^{\infty} a_k$, $\sum_{k=n}^{\infty} a_k$ $(n = 1, 2, 3, \cdots)$, $\sum_{n=1}^{\infty} \prod_{k=n}^{\infty} a_k$, and $\prod_{n=1}^{\infty} \sum_{k=n}^{\infty} a_k$ exist whenever the sequence a_n is bounded. Thus Kantorovitch may say a sequence a_n to converge if it is bounded, and

$$\sum_{n=1}^{\infty} \prod_{k=n}^{\infty} a_k = \prod_{n=1}^{\infty} \sum_{k=n}^{\infty} a_k .$$

In this paper the above convergence-definition will be considerably generalized for the case that the given lattice is a *Boolean ring*. We shall define the convergence of a sequence of elements a_n of a Boolean ring without supposing that $\prod_{k=n}^{\infty} a_k$, $\sum_{k=n}^{\infty} a_k$ $(n = 1, 2, 3, \cdots)$, $\sum_{n=1}^{\infty} \prod_{k=n}^{\infty} a_k$, and $\prod_{n=1}^{\infty} \sum_{k=n}^{\infty} a_k$ exist; we shall not even suppose that the sequence a_n is bounded.

The formulation of the convergence-definition will be the topic of §2 of this paper. In §3 we shall show that in the case that $\prod_{k=n}^{\infty} a_k$ and $\sum_{k=n}^{\infty} a_k$ $(n = 1, 2, 3, \cdots)$ exist, our convergence-definition coincides with the definition given in the introduction. In §4 we shall introduce the important notion of an *invariant* subring $\mathfrak{a}$ of a given Boolean ring A. We shall call $\mathfrak{a}$ an invariant subring of A, or A an invariant extension of $\mathfrak{a}$, if every infinite join of elements of $\mathfrak{a}$, whenever it exists with respect to $\mathfrak{a}$, exists also with respect to A, and has the same value. In §5 we shall define the notion of a *join-extension* A of a Boolean ring $\mathfrak{a}$. A will be said to be a join-extension of $\mathfrak{a}$ if every element of A is a join of elements of $\mathfrak{a}$ with respect to A. We shall see that any join-extension is an invariant extension. The notions of "invariant extension" and "join-extension" will play an important rôle in the following paragraphs. In §6 we shall show that the finite fundamental operations of a Boolean ring are continuous under the introduced sequential topology; moreover we shall deduce a number of other relations. In §7 we shall say a few words about the closure-topology determined by the introduced sequential topology. In §8 we shall state how far the theorems on simple sequences in a Boolean ring obtained hitherto can be extended to double sequences. In §9 we shall, following D. van Dantzig's paper (I), define the notion of a fundamental sequence of elements of a Boolean ring, and explore how far D. van Dantzig's results can be applied to our case. In §10 we shall show that the sequential topology determined by the closure-topology mentioned in §7 is in general not identical with the original sequential topology. I have no further entered into the examination of this derivative sequential topology, but leave it to further investigations.

It is very probable that the theory developed in this paper can be easily extended to the lattices which Fritz Klein in his paper (IV) calls "Boole-Schrödersche Verbände." I shall deal with this matter in a later paper.

The extension of this theory to arbitrary lattices might be more complicated. It is for instance not true that the finite fundamental operations of an arbitrary lattice are continuous with respect to the sequential topology defined in this

introduction. Let a partially ordered set A consist of an enumerated sequence of elements a_n, of another such sequence b_n, and of a further element e, between which there are defined the following ordering relations:

$$a_m < a_n, \qquad b_m < b_n \qquad \text{for } m < n;$$

$$a_m < b_n \qquad \text{for } m \leqq n;$$

$$a_m < e, \qquad b_m < e.$$

$$(m, n = 1, 2, 3, \cdots.)$$

Then A is obviously a complete distributive lattice. ADDED IN PROOF: A is isomorphic with Mac Neille's lattice [4.7, 4.12, 4.7]. See H. M. Mac Neille, *Partially ordered sets*, Trans. Am. Math. Soc., vol. 42 (1937), pp. 416–460, especially Theorem 4.15. If $m > n$, then

$$a_m b_n = a_n,$$

$$a_m \vee b_n = b_m.$$

Obviously we have, by the convergence-definition given in this introduction

$$\lim_{n \to \infty} a_n = e;$$

on the other hand we have for $n = 1, 2, 3, \cdots$

$$a_n b_1 = a_1,$$

whence

$$\lim_{n \to \infty} a_n b_1 = a_1,$$

while

$$e b_1 = b_1.$$

2. The Sequential Topology $\tau(A)$

DEFINITION 1. *By a subelement of a sequence of elements a_n of a Boolean ring A, we mean an element u of A of the property that, if we suitably choose the natural number k, we have for $n \geqq k$*

$$u < a_n.$$

It is clear that the property of an element u of A to be a subelement of the sequence a_n is preserved if we replace a finite number of its terms by other elements, if we add or omit a finite number of terms, or if we change the succession of the terms (not necessarily only of a finite number). The same holds for the concepts of superelement, interelement, lower limit, upper limit, and limit, which we shall define more fully below.

Any sequence of elements of a Boolean ring has at least one subelement, namely the zero-element.

THEOREM 1. *The subelements of a sequence of elements a_n of a Boolean ring A constitute an ideal in A.*

THEOREM 2. *If u is a subelement of a sequence a_n, and b is an arbitrary element, then ub is a subelement of the sequence a_nb, and to every subelement u^* of the sequence a_nb there exists a subelement u of the sequence a_n such that $u^* = ub$.*

DEFINITION 2. *By a superelement of a sequence a_n, we mean an element o of the property that, if we suitably choose the natural number k, we have for $n \geqq k$*

$$o > a_n .$$

It is clear that a sequence of elements of a Boolean ring A is bounded in A if and only if it has at least one superelement in A. If A has a unit every sequence of elements of A is bounded.

THEOREM 3. *The superelements of a bounded sequence of elements of a Boolean ring A constitute a dual ideal in A.* (This dual ideal is a Boole-Schröder lattice in the sense of the paper (IV).)

THEOREM 4. *If o is a superelement of a bounded sequence a_n, and b an arbitrary element, then $o \vee b$ is a superelement of the sequence $a_n \vee b$, and to every superelement o^* of the sequence $a_n \vee b$ there exists a superelement o of the sequence a_n such that $o^* = o \vee b$.*

DEFINITION 3. *Let a_n be a bounded sequence of elements of a Boolean ring A, and also a an element of A. In this case we call a an interelement of the sequence a_n with respect to A, if*

$$u < a < o$$

whenever u and o are elements of A, and u is a subelement, and o a superelement of the sequence a_n.

DEFINITION 4. *Let a_n be an unbounded sequence of elements of a Boolean ring A, and also a an element of A. In this case we call a an interelement of the sequence a_n with respect to A if ab is, according to Definition 3, an interelement of the bounded sequence a_nb with respect to A for every element b of A.*

Instead of saying "a is an interelement of the sequence a_n with respect to A" we may say more simply "a is an interelement of the sequence a_n" if there is no doubt to which Boolean ring the term "interelement" is referred.

THEOREM 5. *If a_n and a are elements of a Boolean ring A, then a is an interelement of the sequence a_n with respect to A if and only if ab is an interelement of the sequence a_nb with respect to A for every element b of A.*

PROOF. As for the case of an unbounded sequence Theorem 5 is involved by Definition 4, we can restrict ourselves to the case that the sequence a_n is bounded.

Let us suppose first that a is an interelement of the bounded sequence a_n, and that a further element b is given. Then Theorem 2 involves that

$$u^* < ab \tag{1}$$

whenever u^* is a subelement of the sequence $a_n b$. If on the other side o^* is a superelement of the sequence $a_n b$, and o a superelement of the given sequence a_n, we have for $n \geqq k$, if k is an appropriate natural number,

$$o^* \vee (o + ob) > a_n b \vee (a_n + a_n b) = a_n ,$$

hence, according to the supposed property of a,

$$o^* \vee (o + ob) > a$$

and

$$o^* > ab. \tag{2}$$

(1) and (2) involve that ab is an interelement of the sequence $a_n b$.

Now let us suppose conversely that ab is an interelement of the sequence $a_n b$ for every element b. From this supposition it yields especially that $a(a \vee c)$ is an interelement of the sequence $a_n(a \vee c)$ whenever c is an upper bound of the sequence a_n. Thus we obtain immediately that a is an interelement of the sequence a_n as we want to prove.

THEOREM 6. *If a_n and a are elements of a Boolean ring A, then a is an interelement of the sequence a_n with respect to A if and only if $a \vee b$ is an interelement of the sequence $a_n \vee b$ with respect to A for every element b of A.*

PROOF. If $a \vee b$ is an interelement of the sequence $a_n \vee b$ for every element b then a is an interelement of the sequence a_n since we may put especially $b = 0$. Now let us suppose conversely that a is an interelement of the sequence a_n, and b an arbitrary element. Let us further suppose that the sequence a_n is *bounded.* In this case the sequence $a_n \vee b$ is bounded too. Let u^* be a subelement, and o^* a superelement of this sequence. Then $u^* + u^*b$ is a subelement of the sequence $a_n + a_n b$, thus also a subelement of the sequence a_n. According to the supposed property of the element a we have

$$u^* + u^*b < a,$$

whence

$$u^* < a \vee b. \tag{3}$$

On the other side Theorem 4 involves the inequality

$$o^* > a \vee b. \tag{4}$$

From (3) and (4) it follows that $a \vee b$ is an interelement of the sequence $a_n \vee b$.

If the sequence a_n is not bounded we observe that ac is, by Definition 4, an interelement of the bounded sequence $a_n c$ for every element c. Hence, according to what we have just proved, $ac \vee bc = (a \vee b)c$ is an interelement of the sequence $a_n c \vee bc = (a_n \vee b)c$. Thus we obtain again the result that $a \vee b$ is an interelement of the sequence $a_n \vee b$.

THEOREM 7. *The interelements of a sequence of elements a_n of a Boolean ring*

A constitute a dense sublattice of A. (This sublattice is a Boole-Schröder lattice in the sense of the paper (IV).)

THEOREM 8. *If u is a subelement, and a an interelement of the same sequence a_n, then $u < a$.*

THEOREM 9. *If a is an interelement of a sequence a_n, and*

$$u < a^* < a$$

whenever u is a subelement of the same sequence, then a^* is as well an interelement of this sequence.

Theorems 7, 8, and 9 follow easily from Theorem 2 and from Definitions 3 and 4. (Notice that $a^{(1)}b < a^{(2)}b$ for every b implies $a^{(1)} < a^{(2)}$.)

DEFINITION 5. *If a_n is a sequence of elements of a Boolean ring A, and the lattice of the interelements of the sequence a_n with respect to A has a zero, this zero is called the lower limit of the sequence a_n with respect to A and denoted by ${}^{(A)}\varliminf_{n\to\infty} a_n$.*

The expression "lower limit of the sequence a_n with respect to A" may be replaced by the simpler expression "lower limit of the sequence a_n" if there is no doubt to which Boolean ring it refers. In this case the sign ${}^{(A)}\varliminf_{n\to\infty} a_n$ may be replaced by the simpler sign $\varliminf_{n\to\infty} a_n$.

If $\varliminf_{n\to\infty} a_n$ exists the lattice of the interelements of the sequence a_n is a Boolean ring, and $\varliminf_{n\to\infty} a_n$ is its zero-element.

LEMMA 1. *If $a > u$ whenever u is a subelement of a given sequence a_n, and $\varliminf_{n\to\infty} a_n$ exists, then*

$$a > \varliminf_{n\to\infty} a_n.$$

PROOF. If u is a subelement of the sequence a_n then from the hypothesis and from Theorem 8 follows that

$$a \varliminf_{n\to\infty} a_n > u.$$

This inequality implies, by Theorem 9, that $a \varliminf_{n\to\infty} a_n$ is an interelement of the sequence a_n. Hence we have, according to Definition 5,

$$a \varliminf_{n\to\infty} a_n > \varliminf_{n\to\infty} a_n$$

or equivalently

$$a > \varliminf_{n\to\infty} a_n$$

as we wished to prove.

DEFINITION 6. *If a_n is a sequence of elements of a Boolean ring A, and the lattice of the interelements of the sequence a_n with respect to A has a unit, this unit is called the upper limit of the sequence a_n with respect to A, and denoted by ${}^{(A)}\varlimsup_{n\to\infty} a_n$.*

The expression "upper limit of the sequence a_n with respect to A" may be replaced by the simpler expression "upper limit of the sequence a_n" if there is no doubt to which Boolean ring it refers. In this case the sign ${}^{(A)}\overline{\lim}_{n\to\infty} a_n$ may be replaced by the simpler sign $\overline{\lim}_{n\to\infty} a_n$.

LEMMA 2. *If $a < o$ whenever o is a superelement of a given bounded sequence a_n, and $\overline{\lim}_{n\to\infty} a_n$ exists, then*

$$a < \overline{\lim_{n\to\infty}}\, a_n .$$

PROOF. If o is a superelement of the sequence a_n then from the hypothesis follows that

$$a \vee \overline{\lim_{n\to\infty}}\, a_n < o.$$

Hence $a \vee \overline{\lim}_{n\to\infty} a_n$ is, by Definition 3, an interelement of the sequence a_n. On the other side we have, by Definition 6,

$$a \vee \overline{\lim_{n\to\infty}}\, a_n < \overline{\lim_{n\to\infty}}\, a_n$$

or equivalently

$$a < \overline{\lim_{n\to\infty}}\, a_n$$

as we wished to prove.

THEOREM 10. *If both $\underline{\lim}_{n\to\infty} a_n$ and $\overline{\lim}_{n\to\infty} a_n$ exist then*

$$\underline{\lim_{n\to\infty}}\, a_n < \overline{\lim_{n\to\infty}}\, a_n .$$

THEOREM 11. *If both $\underline{\lim}_{n\to\infty} a_n$ and $\overline{\lim}_{n\to\infty} a_n$ exist then a is an interelement of the sequence a_n if and only if*

$$\underline{\lim_{n\to\infty}}\, a_n < a < \overline{\lim_{n\to\infty}}\, a_n .$$

We need now some transfinite extensions of the known rules which the finite operations of a Boolean ring obey.

THEOREM 12. *Let $\mathfrak{B}$ be a class of elements of a Boolean ring A, and also b an element of A. If in this case $\prod_{a\in\mathfrak{B}} a$ exists then $\prod_{a\in\mathfrak{B}} (a \vee b)$ exists too and is equal to $(\prod_{a\in\mathfrak{B}} a) \vee b$.*

PROOF. It is obviously sufficient to prove that

$$c < a \vee b \tag{5}$$

for every element a of $\mathfrak{B}$ implies

$$c < (\prod_{a\in\mathfrak{B}} a) \vee b.$$

From (5) follows that

$$c + cb < a$$

for every element a of $\mathfrak{B}$. Hence

$$c + cb < \prod_{a \in \mathfrak{B}} a,$$

and

$$c < (\prod_{a \in \mathfrak{B}} a) \vee b$$

as we wished to prove.

In the following Theorems 13 till 17 the signs $\mathfrak{B}$ and b shall have the same significance as in Theorem 12. In the special case of Boolean σ-rings they were proved by J. von Neumann, "Lectures on continuous geometries," Princeton, 1937, Appendix, p. 7. See also H. M. Mac Neille, l.c., Theorem 7.19.

THEOREM 13. *If* $\sum_{a \in \mathfrak{B}} a$ *exists then* $\sum_{a \in \mathfrak{B}} ab$ *exists too and is equal to* $(\sum_{a \in \mathfrak{B}} a)b$.

PROOF. If $c > ab$ for every element a of $\mathfrak{B}$ then

$$c \vee [\sum_{x \in \mathfrak{B}} x + (\sum_{x \in \mathfrak{B}} x)b] > ab \vee [\sum_{x \in \mathfrak{B}} x + (\sum_{x \in \mathfrak{B}} x)b] > a \text{ for } a \in \mathfrak{B};$$

hence

$$c \vee [\sum_{x \in \mathfrak{B}} x + (\sum_{x \in \mathfrak{B}} x)b] > \sum_{a \in \mathfrak{B}} a,$$

and

$$c > (\sum_{a \in \mathfrak{B}} a)\, b.$$

THEOREM 14. *If* $\prod_{a \in \mathfrak{B}} a$ *exists then* $\sum_{a \in \mathfrak{B}} (b + ba)$ *exists too and is equal to* $b + b \prod_{a \in \mathfrak{B}} a$.

PROOF. It is clear that

$$b + b \prod_{a \in \mathfrak{B}} a > b + bx$$

for $x \in \mathfrak{B}$. If $c > b + ba$ for every element a of $\mathfrak{B}$ then

$$a > b + bc$$

for $a \in \mathfrak{B}$; hence

$$\prod_{a \in \mathfrak{B}} a > b + bc,$$

and

$$c > b + b \prod_{a \in \mathfrak{B}} a.$$

THEOREM 15. *If $\sum_{a \in \mathfrak{B}} a$ exists then $\prod_{a \in \mathfrak{B}} (b + ba)$ exists too and is equal to $b + b\sum_{a \in \mathfrak{B}} a$.*

PROOF. If $c < b + ba$ for $a \in \mathfrak{B}$ then $c < b$ and $ca = 0$ for $a \in \mathfrak{B}$; thus, by Theorem 13, $c\sum_{a \in \mathfrak{B}} a = 0$, and $c < b + b\sum_{a \in \mathfrak{B}} a$.

THEOREM 16. *If $\prod_{a \in \mathfrak{B}} a$ exists then $\prod_{a \in \mathfrak{B}} (a + ab)$ exists too and is equal to $\prod_{a \in \mathfrak{B}} a + (\prod_{a \in \mathfrak{B}} a)b$.*

PROOF. If $c < a + ab$ for $a \in \mathfrak{B}$ then $c < a$ for $a \in \mathfrak{B}$, thus $c < \prod_{a \in \mathfrak{B}} a$; on the other side $cb = 0$; hence

$$c < \prod_{a \in \mathfrak{B}} a + (\prod_{a \in \mathfrak{B}} a)\, b.$$

THEOREM 17. *If $\sum_{a \in \mathfrak{B}} a$ exists then $\sum_{a \in \mathfrak{B}} (a + ab)$ exists too and is equal to $\sum_{a \in \mathfrak{B}} a + (\sum_{a \in \mathfrak{B}} a)b$.*

PROOF. If $c > a + ab$ for $a \in \mathfrak{B}$ then $c \vee b > a$ for $a \in \mathfrak{B}$, thus $c \vee b > \sum_{a \in \mathfrak{B}} a$, and $c > \sum_{a \in \mathfrak{B}} a + (\sum_{a \in \mathfrak{B}} a)b$.

THEOREM 18. *If $\mathfrak{a}$ is an ideal in a Boolean ring A, then $\mathfrak{a}''$ is the class of all elements b of A such that $b = \sum_{a \in \mathfrak{B}}^{(A)} a$ for an appropriate subclass $\mathfrak{B}$ of $\mathfrak{a}$. $\mathfrak{a}''$ is a principal ideal if and only if $\sum_{a \in \mathfrak{a}}^{(A)} a$ exists; in this case $\mathfrak{a}'' = \mathfrak{a}(\sum_{a \in \mathfrak{a}}^{(A)} a)$.* ADDED IN PROOF: Theorem 18 is partly contained in Theorem 2.1 of M. H. Stone's paper on *Algebraic characterizations of special Boolean rings*, Fundamenta Mathematicae, vol. 29 (1937), pp. 223–303.

PROOF. From the definition of $\mathfrak{a}''$ and from Theorem 13 follows that $\mathfrak{a}''$ contains the element $\sum_{a \in \mathfrak{B}}^{A} a$ whenever $\mathfrak{B} \subset \mathfrak{a}$, and $\sum_{a \in \mathfrak{B}}^{A} a$ exists. If b is an arbitrary element of $\mathfrak{a}''$ we put $\mathfrak{B} = \mathfrak{a} \cdot \mathfrak{a}(b)$. Then $b > a$ whenever a is an element of $\mathfrak{B}$. If c is an arbitrary element of A then

$$(b + bc)x \in \mathfrak{B}$$

for all elements x of $\mathfrak{a}$. Hence if $c \in A$ and $c > a$ for every element a of $\mathfrak{B}$ we have, for $x \in \mathfrak{a}$,

$$(b + bc)x < c,$$

thus

$$(b + bc)x = 0$$

or

$$b + bc \in \mathfrak{a}'.$$

Since $b \in \mathfrak{a}''$ we have $b + bc = 0$, thus $b < c$. Hence

$$b = \sum_{a \in \mathfrak{B}}^{(A)} a.$$

This proves the first part of the theorem. The second part is now obvious.

THEOREM 19. *$\underline{\lim}_{n\to\infty} a_n$ exists if and only if $\sum_{u \in \mathfrak{U}} u$, $\mathfrak{U}$ being the class of the subelements of the sequence a_n, exists; in this case*

$$\underline{\lim}_{n\to\infty} a_n = \sum_{u \in \mathfrak{U}} u. \tag{6}$$

PROOF. If $\underline{\lim}_{n\to\infty} a_n$ exists, from Lemma 1 and Theorem 8 follows immediately that $\sum_{u \in \mathfrak{U}} u$ exists, and the equation (6) holds. If $\sum_{u \in \mathfrak{U}} u$ exists then there exists also $\sum_{u \in \mathfrak{U}} ub$ for every element b. (See Theorem 13.) If we notice Theorem 2, we conclude easily that $\sum_{u \in \mathfrak{U}} ub$ is an interelement of the bounded sequence $a_n b$. On the other side we have

$$\sum_{u \in \mathfrak{U}} ub = \Big(\sum_{u \in \mathfrak{U}} u\Big) b.$$

Hence $\sum_{u \in \mathfrak{U}} u$ is an interelement of the sequence a_n. If a is an arbitrary interelement of this sequence, we have, by Theorem 8,

$$a > u$$

for every element u of $\mathfrak{U}$, and consequently

$$a > \sum_{u \in \mathfrak{U}} u.$$

Hence $\underline{\lim}_{n\to\infty} a_n$ exists and is equal to $\sum_{u \in \mathfrak{U}} u$.

THEOREM 20. *$\underline{\lim}_{n\to\infty} a_n$ exists if and only if $\mathfrak{U}''$, $\mathfrak{U}$ having the same significance as in Theorem 19, is a principal ideal $\mathfrak{a}(a)$; in this case*

$$\underline{\lim}_{n\to\infty} a_n = a.$$

PROOF. See Theorems 18 and 19.

THEOREM 21. (Corollary of Theorem 20.) *If the ideal of the subelements of a sequence a_n is a principal ideal $\mathfrak{a}(a)$, then $\underline{\lim}_{n\to\infty} a_n$ exists and is equal to a.*

THEOREM 22. *The upper limit of a bounded sequence a_n exists if and only if $\prod_{o \in \mathfrak{O}} o$, $\mathfrak{O}$ being the class of the superelements of the sequence a_n, exists, in this case*

$$\overline{\lim}_{n\to\infty} a_n = \prod_{o \in \mathfrak{O}} o.$$

The proof can be left to the reader. (Notice Lemma 2.)

THEOREM 23. (Corollary of Theorem 22.) *If the dual ideal of the superelements of a bounded sequence a_n has a zero then $\overline{\lim}_{n\to\infty} a_n$ exists and is equal to this zero.*

In this case the dual ideal of the superelements of the sequence a_n is a Boolean ring, and $\overline{\lim}\, a_n$ is its zero.

LEMMA 3. *If* $\underline{\lim}_{n\to\infty} a_n$ *exists, and b is an arbitrary element, then* $\underline{\lim}_{n\to\infty} a_n b$ *exists too and is equal to b* $\underline{\lim}_{n\to\infty} a_n$.

PROOF. See Theorems 2, 13, and 19.

LEMMA 4. *If* $\overline{\lim}_{n\to\infty} a_n$ *exists, and b is an arbitrary element, then* $\overline{\lim}_{n\to\infty} a_n b$ *exists too and is equal to b* $\overline{\lim}_{n\to\infty} a_n$.

PROOF. Let $\overline{\lim}_{n\to\infty} a_n$ be denoted by a, and let c be an interelement of the sequence $a_n b$. Further let d be an arbitrary element, and o a superelement of the sequence $a_n d$. Thus o is also a superelement of the sequence $a_n bd$, and $cd < o$. We have also $ad < o$ because ad is an interelement of the sequence $a_n d$. Hence

$$(a \vee c)d < o.$$

If on the other side u is a subelement of the sequence $a_n d$, then $ad > u$, whence

$$(a \vee c)d > u.$$

We find that $a \vee c$ is an interelement of the sequence a_n . This involves, by Definition 6, $a \vee c < a$, or equivalently $c < a$. Besides we have obviously $c < b$. Hence $c < ab$ or

$$c < b \overline{\lim_{n\to\infty}} a_n.$$

Since $b \overline{\lim}_{n\to\infty} a_n$ is itself an interelement of the sequence $a_n b$, Lemma 4 is proved.

DEFINITION 7. *If a sequence of elements* a_n *of a Boolean ring A has exactly one interelement a with respect A, we say that the sequence* a_n *converges towards a with respect to A, or a is its limit with respect to A, and write*

$$^{(A)}\lim_{n\to\infty} a_n = a.$$

The sentences "The sequence a_n converges towards a with respect to A" and "a is the limit of the sequence a_n with respect to A" may be replaced respectively by the shorter sentences "The sequence a_n converges towards a" and "a is the limit of the sequence a_n" if there is no doubt to which Boolean ring they refer. In this case the sign $^{(A)}\lim_{n\to\infty} a_n$ may be replaced by the simpler sign $\lim_{n\to\infty} a_n$.

DEFINITION 8. *The sequential topology in a Boolean ring A, introduced by Definition 7, shall be denoted by* $\tau(A)$.

THEOREM 24. *A sequence of elements* a_n *of a Boolean ring is convergent if and only if* $\underline{\lim}_{n\to\infty} a_n$ *and* $\overline{\lim}_{n\to\infty} a_n$ *exist and are equal to one another. In this case*

$$\lim_{n\to\infty} a_n = \underline{\lim}_{n\to\infty} a_n = \overline{\lim}_{n\to\infty} a_n .$$

LEMMA 5. *If* $\lim_{n\to\infty} a_n$ *exists, and b is an arbitrary element, then* $\lim_{n\to\infty} a_n b$ *exists too and is equal to b* $\lim a_n$.

PROOF. See Lemmas 3 and 4 and Theorem 24.

THEOREM 25. *If $a_n = a$ for every n then also*

$$\lim_{n \to \infty} a_n = a.$$

THEOREM 26. *A subelement (superelement) of a given sequence is also a subelement (superelement) of any partial sequence.*

THEOREM 27. *Any interelement of a partial sequence of a given sequence is also an interelement of the whole sequence.*

Theorem 26 is evident, Theorem 27 can be easily deduced from Theorem 26.

THEOREM 28. *If a_n is an arbitrary sequence of elements of a Boolean ring, and n_k $(k = 1, 2, 3, \cdots)$ are natural numbers different from each other, then there hold always those of the inclusion-relations*

$$\underline{\lim}_{k \to \infty} a_{n_k} > \underline{\lim}_{n \to \infty} a_n,$$

$$\underline{\lim}_{k \to \infty} a_{n_k} < \overline{\lim}_{n \to \infty} a_n,$$

$$\overline{\lim}_{k \to \infty} a_{n_k} > \underline{\lim}_{n \to \infty} a_n,$$

and

$$\overline{\lim}_{k \to \infty} a_{n_k} < \overline{\lim}_{n \to \infty} a_n,$$

in which both sides exist.

PROOF. See Definitions 5 and 6 and Theorem 27.

THEOREM 29. *If $\lim_{n \to \infty} a_n$ exists, and n_k $(k = 1, 2, 3, \cdots)$ are natural numbers different from each other, then $\lim_{k \to \infty} a_{n_k}$ exists too and is equal to $\lim_{n \to \infty} a_n$.*

PROOF. Let us suppose first that the sequence a_n is bounded. Then, if u is a subelement and o a superelement of this sequence, and u^* and o^* have the same meaning for the sequence a_{n_k} $(k = 1, 2, 3, \cdots)$, we have, by Theorem 26,

$$u^* < o$$

and

$$u < o^*.$$

Hence we have, by Lemmas 1 and 2,

$$u^* < \overline{\lim}_{n \to \infty} a_n$$

and

$$\underline{\lim}_{n \to \infty} a_n < o^*,$$

i.e.

$$u^* < \lim_{n \to \infty} a_n < o^*.$$

We see that $\lim a_n$ is an interelement of the sequence a_{n_k} $(k = 1, 2, 3, \cdots)$.

The same result is obtained also in the case that the sequence a_n is not bounded. For if b is an arbitrary element Lemma 5 yields that $\lim_{n\to\infty} a_n b$ exists and is equal to $b \lim_{n\to\infty} a_n$; hence $b \lim_{n\to\infty} a_n$ is, by what we have just proved, an interelement of the sequence $a_{n_k} b$ $(k = 1, 2, 3, \cdots)$.

$\lim_{n\to\infty} a_n$ is also the only interelement of the sequence a_{n_k} $(k = 1, 2, 3, \cdots)$; for if the sequence a_{n_k} had an interelement a different from $\lim_{n\to\infty} a_n$ then a would be, by Theorem 27, also an interelement of the original sequence a_n , contrary to the hypothesis that this sequence is convergent. By this Theorem 29 is proved.

It can occur that both $\underline{\lim}_{n\to\infty} a_n$ and $\overline{\lim}_{n\to\infty} a_n$ exist but that there is neither a partial sequence of the sequence a_n converging towards $\underline{\lim}_{n\to\infty} a_n$ nor such a sequence converging towards $\overline{\lim}_{n\to\infty} a_n$. Let a and b be two elements such that neither $a < b$ nor $b < a$. Now for $n = 1, 2, 3, \cdots$ let $a_{2n-1} = a$, $a_{2n} = b$. Then we have obviously

$$\underline{\lim}_{n\to\infty} a_n = ab$$

and

$$\overline{\lim}_{n\to\infty} a_n = a \vee b.$$

Nevertheless there is neither a partial sequence of the given sequence converging towards ab nor such a sequence converging towards $a \vee b$.

THEOREM 30. *The union of a finite number of convergent sequences with the same limit is again a convergent sequence with the same limit.*

PROOF. Suppose that $a_n^{(1)}$ and $a_n^{(2)}$ are two convergent sequences, and that

$$\lim_{n\to\infty} a_n^{(1)} = \lim_{n\to\infty} a_n^{(2)} = a.$$

Suppose moreover that the sequence a_n consists of the terms of the sequences $a_n^{(1)}$ and $a_n^{(2)}$ in some succession. Then from Theorem 27 follows that a is an interelement of the sequence a_n . Now let a^* be an arbitrary interelement of this sequence, b an arbitrary element, $u^{(1)}$ and $u^{(2)}$ subelements, and $o^{(1)}$ and $o^{(2)}$ superelements of the sequences $a_n^{(1)}b$ and $a_n^{(2)}b$ respectively. Then $u^{(1)}u^{(2)}$ is a subelement, and $o^{(1)} \vee o^{(2)}$ a superelement of the sequence a_n . Since a^*b is an interelement of this sequence we have

$$u^{(1)}u^{(2)} \dot{<} a^*b < o^{(1)} \vee o^{(2)}. \tag{7}$$

The sequences $a_n^{(1)}b$ and $a_n^{(2)}b$ are convergent, and

$$\lim_{n\to\infty} a_n^{(1)}\, b = \lim_{n\to\infty} a_n^{(2)}\, b = ab.$$

(See Lemma 5.) Hence we get from (7) by using twice Theorem 19 together with Theorem 13, and Theorem 22 together with Theorem 12

$$ab \cdot ab < a^*b < ab \vee ab$$

i.e.

$$a^*b = ab.$$

Since b is arbitrary we obtain the result $a^* = a$. This proves the theorem.

3. Connection with the Usual Convergence-Definition

The following results show that our definition specializes in the case of σ-lattices to the usual definition (cf. for example G. Birkhoff, II, p. 453).

THEOREM 31. *If $a_k > a_l$ whenever $k < l$, the sequence a_n is either convergent or has no interelement; it is convergent if and only if $\prod_{k=1}^{\infty} a_k$ exists; in this case we have*

$$\prod_{k=1}^{\infty} a_k = \lim_{n \to \infty} a_n . \tag{8}$$

PROOF. If a is an interelement of the sequence a_n, we have

$$a < a_n$$

for every n because all terms of the sequence are superelements of the sequence. If $u < a_n$ for every n then u is a subelement of the sequence, and we have

$$a > u.$$

Hence $\prod_{k=1}^{\infty} a_k$ exists and is equal to a. Thus we see that a is the only interelement of our sequence, is to be denoted by $\lim_{n \to \infty} a_n$, and the equation (8) holds.

If conversely $\prod_{k=1}^{\infty} a_k$ exists then it can be easily concluded that

$$u < \prod_{k=1}^{\infty} a_k < o$$

whenever u is a subelement and o a superelement of the sequence a_n, so that $\prod_{k=1}^{\infty} a_k$ is an interelement of this sequence.

THEOREM 32. (Corollary of Theorem 31.) *The sequence $\prod_{k=1}^{n} a_k$ ($n = 1, 2, 3, \cdots$) is either convergent or has no interelement; it is convergent if and only if $\prod_{k=1}^{\infty} a_k$ exists; in this case we have*

$$\prod_{k=1}^{\infty} a_k = \lim_{n \to \infty} \prod_{k=1}^{n} a_k .$$

THEOREM 33. *If $a_k < a_l$ whenever $k < l$, the sequence a_n is either convergent or has no interelement; it is convergent if and only if $\sum_{k=1}^{\infty} a_k$ exists; in this case we have*

$$\sum_{k=1}^{\infty} a_k = \lim_{n \to \infty} a_n .$$

If the suppositions of Theorem 33 are satisfied, and the sequence a_n has an interelement a, then we have

$$a > a_n$$

for every n because all terms of the sequence are subelements of the sequence. Hence the sequence a_n is certainly bounded. The same is also true if $\sum_{k=1}^{\infty} a_k$ exists. After this is settled Theorem 33 can be proved in a similar way as Theorem 31.

THEOREM 34. (Corollary of Theorem 33.) *The sequence* $\sum_{k=1}^{n^*} a_k$ $(n = 1, 2, 3, \cdots)$ *is either convergent or has no interelement; it is convergent if and only if* $\sum_{k=1}^{\infty} a_k$ *exists; in this case*

$$\sum_{k=1}^{\infty} a_k = \lim_{n\to\infty} \sum_{k=1}^{n} a_k .$$

THEOREM 35. *Any monotonic sequence of elements of a Boolean σ-ring is convergent.*

Theorem 35 follows immediately from Theorems 31 and 33.

THEOREM 36. *If for* $n = 1, 2, 3, \cdots$ $\prod_{k=n}^{\infty} a_k$ *exists then* $\underline{\lim}_{n\to\infty} a_n$ *exists if and only if* $\sum_{n=1}^{\infty} \prod_{k=n}^{\infty} a_k$ *exists; in this case we have*

$$\underline{\lim}_{n\to\infty} a_n = \sum_{n=1}^{\infty} \prod_{k=n}^{\infty} a_k = \lim_{n\to\infty} \prod_{k=n}^{\infty} a_k .$$

PROOF. Definition 1 involves that to every subelement u of the sequence a_n there can be stated a natural number m so that

$$\prod_{k=m}^{\infty} a_k > u.$$

From this it can be easily concluded that $\sum_{u \in \mathfrak{U}} u$ exists if and only if $\sum_{n=1}^{\infty} \prod_{k=n}^{\infty} a_k$ exists, and that in this case both elements are identical. ($\mathfrak{U}$ is the ideal of the subelements of the sequence a_n.) Now Theorems 19 and 33 show that Theorem 36 is true.

THEOREM 37. *If for* $n = 1, 2, 3, \cdots$ $\sum_{k=n}^{\infty} a_k$ *exists then* $\overline{\lim}_{n\to\infty} a_n$ *exists if and only if* $\prod_{n=1}^{\infty} \sum_{k=n}^{\infty} a_k$ *exists; in this case we have*

$$\overline{\lim}_{n\to\infty} a_n = \prod_{n=1}^{\infty} \sum_{k=n}^{\infty} a_k = \lim_{n\to\infty} \sum_{k=n}^{\infty} a_k .$$

Theorem 37 can be proved by the aid of Theorems 22 and 31 in a similar way as Theorem 36 by the aid of Theorems 19 and 33.

THEOREM 38. *If a_n is a sequence of elements of a Boolean σ-ring then there exist certainly both* $\underline{\lim}_{n\to\infty} a_n$ *and* $\overline{\lim}_{n\to\infty} a_n$.

PROOF. See Theorems 36 and 37.

THEOREM 39. *If* a_n *is a sequence of elements of a Boolean* σ*-ring then* $\lim_{n\to\infty} a_n$ *exists if and only if*

$$\sum_{n=1}^{\infty}\prod_{k=n}^{\infty} a_k = \prod_{n=1}^{\infty}\sum_{k=n}^{\infty} a_k .$$

In this case we have

$$\lim_{n\to\infty} a_n = \sum_{n=1}^{\infty}\prod_{k=n}^{\infty} a_k = \prod_{n=1}^{\infty}\sum_{k=n}^{\infty} a_k .$$

Thus our Definition 7 coincides in the case of a Boolean σ-ring with the definition of convergence and limit given in the introduction.

LEMMA 6. *If* a_n *is a sequence of elements of a Boolean* σ*-ring with unit then*

$$\varliminf_{n\to\infty} a_n' = (\varlimsup_{n\to\infty} a_n)',$$

$$\varlimsup_{n\to\infty} a_n' = (\varliminf_{n\to\infty} a_n)';$$

if moreover the sequence a_n *is convergent, the sequence* a_n' *is convergent too, and*

$$\lim_{n\to\infty} a_n' = (\lim_{n\to\infty} a_n)'.$$

Lemma 6 can be easily deduced from Theorems 14, 15, 36, and 37.

LEMMA 7. *If* a_n *and* a *are elements of a Boolean* σ*-ring with unit, and* a *is an interelement of the sequence* a_n , *then* a' *is an interelement of the sequence* a_n' .

PROOF. See Theorem 11 and Lemma 6.

THEOREM 40. *If* $a_n > c$ *for every* n *then the equation*

$$\varliminf_{n\to\infty} a_n = c \tag{9}$$

holds if and only if for every n

$$\prod_{k=n}^{\infty} a_k = c. \tag{10}$$

PROOF. If (9) holds, and the element u is of the property that for $k \geqq n$ $u < a_k$, then $u < c$ because u is a subelement of the sequence a_n . This proves (10). If (10) holds then (9) follows immediately from Theorem 36.

THEOREM 41. *If* $a_n < c$ *for every* n *then the equation*

$$\varlimsup_{n\to\infty} a_n = c$$

holds if and only if for every n

$$\sum_{k=n}^{\infty} a_k = c.$$

THEOREM 42. *An element a is an interelement of a sequence a_n if and only if for every n*

$$\sum_{k=n}^{\infty} a_k a = a, \tag{11}$$

and

$$\prod_{k=n}^{\infty} (a_k \vee a) = a. \tag{12}$$

PROOF. If a is an interelement of the sequence a_n then it is, by Theorems 5 and 6, also an interelement of the sequences $a_n a$ and $a_n \vee a$. On the other side it is a superelement of the sequence $a_n a$ and a subelement of the sequence $a_n \vee a$. Hence we have

$$\overline{\lim_{n \to \infty}}\, a_n a = a,$$

$$\underline{\lim_{n \to \infty}}\, (a_n \vee a) = a.$$

Now Theorems 41 and 40 involve that the equations (11) and (12) hold.

Let us suppose conversely that the equations (11) and (12) hold. Let b be an arbitrary element, and u a subelement and o a superelement of the sequence $a_n b$. Further let m be a natural number of the property that

$$u < a_k b < o$$

for $k \geqq m$. Then we have

$$ab = \sum_{k=m}^{\infty} a_k ab < ao,$$

whence

$$ab < o.$$

Similarly we have

$$a = \prod_{k=m}^{\infty} (a_k \vee a) > u \vee a,$$

whence

$$u < a.$$

Since also $u < b$ we get

$$u < ab.$$

Thus we see that ab is an interelement of the sequence $a_n b$; hence a is an interelement of the sequence a_n as we wished to prove.

Another form of Theorem 42 is

THEOREM 43. *An element a is an interelement of a sequence a_n if and only if*

$$\overline{\lim_{n\to\infty}}\, a_n a = a, \tag{13}$$

and

$$\underline{\lim_{n\to\infty}}\, (a_n \vee a) = a. \tag{14}$$

PROOF. See Theorems 40, 41 and 42.

The following Theorems 44 and 45 will be needed in the next paragraph.

THEOREM 44. *The equation*

$$\underline{\lim_{n\to\infty}}\, a_n = a \tag{15}$$

is equivalent with the couple of the equations

$$\lim_{n\to\infty} a_n a = a \tag{16}$$

and

$$\underline{\lim_{n\to\infty}}\, (a_n \vee a) = a. \tag{14}$$

PROOF. That (15) implies (14) and (16) follows from Lemma 3 and Theorem 43. If (14) and (16) hold then a is, by Theorem 43, an interelement of the sequence a_n; if b is another such interelement then, by Theorem 5 and (16), $ab = a$, i.e. $a < b$.

THEOREM 45. *The equation*

$$\overline{\lim_{n\to\infty}}\, a_n = a \tag{17}$$

is equivalent with the couple of the equations

$$\overline{\lim_{n\to\infty}}\, a_n a = a \tag{13}$$

and

$$\lim_{n\to\infty} (a_n \vee a) = a. \tag{18}$$

PROOF. If (17) holds we have to prove, in face of Theorem 43, the equation (18) only. Let b be an arbitrary element, and o a superelement of the sequence $a_n b$. Then we have $o > ab$. If moreover c is an interelement of the sequence $a_n \vee a$ then bc is an interelement of the sequence $a_n b \vee ab$; since o is a superelement also of this sequence we have $bc < o$; thus, by Lemmas 2 and 4, $bc < ab$. Since b is arbitrary we have $c < a$. On the other side we have certainly $c > a$. Hence $c = a$. This proves (18).

If (13) and (18) hold then a is, by Theorem 43, an interelement of the sequence a_n; if b is another such interelement then, by Theorem 6 and (18), $a \vee b = a$, thus $a > b$. This proves (17).

4. Invariant subrings

If $\mathfrak{a}$ is a subring of a Boolean ring A, the restriction of the sequential topology $\tau(A)$ (Definition 8) to the subring $\mathfrak{a}$ is not necessarily identical with $\tau(\mathfrak{a})$.

Let for instance A be the Boolean algebra of all classes of natural numbers, a_n the class of all natural numbers divisible by n ($n = 1, 2, 3, \cdots$), and $\mathfrak{a}$ the subring of A generated by the elements a_n ($n = 1, 2, 3, \cdots$) of A. $\mathfrak{a}$ contains the unit of A. If a is a polynomial in the elements a_k ($k = 1, 2, \cdots, m$), and M the least common multiple of the natural numbers $1, 2, 3, \cdots, m$, then a natural number l is an element of a if and only if $l + M$ is an element of a. This yields that every element of $\mathfrak{a}$ is either an infinite class or the void class.

Now the classes $\prod_{k=n}^{\infty}{}^{(A)} a_k' = \left(\sum_{k=n}^{\infty}{}^{(A)} a_k\right)'$ are finite. Hence we have

$$\prod_{k=n}^{\infty}{}^{(\mathfrak{a})} a_k' = 0$$

and

$$\sum_{k=n}^{\infty}{}^{(\mathfrak{a})} a_k = e$$

if e is the unit of A. From the last equation follows

$$^{(\mathfrak{a})}\overline{\lim_{n\to\infty}}\, a_n = e.$$

It can also easily be proved that

$$^{(\mathfrak{a})}\underline{\lim_{n\to\infty}}\, a_n = 0.$$

We see that the sequence a_n is not convergent with respect to $\mathfrak{a}$. On the other hand we have obviously

$$^{(A)}\lim_{n\to\infty} a_n = 0.$$

Thus $\tau(\mathfrak{a})$ is not identical with the restriction of $\tau(A)$ to the subring $\mathfrak{a}$.

THEOREM 46. *If $\mathfrak{a}$ is a subring of a Boolean ring A, a_n ($n = 1, 2, 3, \cdots$) and a are elements of $\mathfrak{a}$, and a is an interelement of the sequence a_n with respect to A, then a is also an interelement of the sequence a_n with respect to $\mathfrak{a}$.*

PROOF. If b is an element of $\mathfrak{a}$ then ab is, by Theorem 5, an interelement of the sequence a_nb with respect to A. From Definition 3 it can be easily concluded that ab is also an interelement of the sequence a_nb with respect to $\mathfrak{a}$. Since this holds for every element b of $\mathfrak{a}$ Theorem 46 is proved.

THEOREM 47. *If $\mathfrak{a}$ is a subring of a Boolean ring A, $\mathfrak{B}$ a subclass of $\mathfrak{a}$, and b an element of $\mathfrak{a}$, then*

$$\prod_{a \in \mathfrak{B}}{}^{(A)} a = b$$

implies

$$\prod_{a \in \mathfrak{B}}^{(\mathfrak{a})} = b,$$

and

$$\sum_{a \in \mathfrak{B}}^{(A)} a = b$$

implies

$$\sum_{a \in \mathfrak{B}}^{(\mathfrak{a})} a = b.$$

DEFINITION 9. *A subring* $\mathfrak{a}$ *of a Boolean ring* A *is said to be invariant in* A *(*A *an invariant extension of* $\mathfrak{a}$*,* A *invariant over* $\mathfrak{a}$*) if the following condition is satisfied: if* $\mathfrak{B}$ *is a subclass of* $\mathfrak{a}$*, and* b *an element of* $\mathfrak{a}$*, then*

$$\sum_{a \in \mathfrak{B}}^{(\mathfrak{a})} a = b$$

implies

$$\sum_{a \in \mathfrak{B}}^{(A)} a = b.$$

Let for instance A be the Boolean algebra of all sets of real numbers and $\mathfrak{a}$ the subclass of A defined in the following way: a is an element of $\mathfrak{a}$ if and only if it is either the empty set or the union of a finite number of sets, every one of which is either an open real interval or consists of exactly one real number. Then it is easily to be seen that $\mathfrak{a}$ is an invariant subring of A.

Let, on the other hand, $\mathfrak{b}$ be the subclass of A defined in the following way: b is an element of $\mathfrak{b}$ if and only if it is either the empty set or the union of a finite number of right-hand open (and left-hand closed) real intervals. Then $\mathfrak{b}$ is as well a subring of A, but is not invariant in A. For if $c(\mu)$ (μ a positive real number) is the set of those real numbers ν which satisfy the inequality $0 \leqq \nu < \mu$ then $\prod_{\mu>0}^{(A)} c(\mu)$ is the set consisting only of the number zero while $\prod_{\mu>0}^{(\mathfrak{b})} c(\mu)$ is the empty set. (See the following Theorem 48.)

THEOREM 48. *If* $\mathfrak{a}$ *is an invariant subring of a Boolean ring* A*,* $\mathfrak{B}$ *a subclass of* $\mathfrak{a}$*, and* b *an element of* $\mathfrak{a}$*, then*

$$\prod_{a \in \mathfrak{B}}^{(\mathfrak{a})} a = b$$

implies

$$\prod_{a \in \mathfrak{B}}^{(A)} a = b.$$

PROOF. If a_0 is a fixed element of $\mathfrak{B}$, we have, by Theorem 14,

$$\sum_{a \in \mathfrak{B}}^{(\mathfrak{a})} (a_0 + a_0 a) = a_0 + a_0 \prod_{a \in \mathfrak{B}}^{(\mathfrak{a})} a = a_0 + b;$$

hence, by Definition 9,

$$\sum_{a \in \mathfrak{B}}^{(A)} (a_0 + a_0 a) = a_0 + b$$

and, by Theorem 15,

$$\prod_{a \in \mathfrak{B}}^{(A)} [a_0 + a_0(a_0 + a_0 a)] = a_0 + a_0(a_0 + b)$$

i.e.

$$\prod_{a \in \mathfrak{B}}^{(A)} a = b$$

as we wished to prove.

THEOREM 49. *If* $\mathfrak{a}$ *is an invariant subring of a Boolean ring* A, a_n $(n = 1, 2, 3, \cdots)$ *and* a *are elements of* $\mathfrak{a}$, *and* a *is an interelement of the sequence* a_n *with respect to* $\mathfrak{a}$, *then* a *is also an interelement of the sequence* a_n *with respect to* A.

PROOF. See Theorem 42, Definition 9, and Theorem 48.

THEOREM 50. *If* $\mathfrak{a}$ *is an invariant subring of a Boolean ring* A, *and* a_n $(n = 1, 2, 3, \cdots)$ *and* a *are elements of* $\mathfrak{a}$, *then the equation*

$$\text{(19)} \qquad {}^{(\mathfrak{a})}\lim_{n \to \infty} a_n = a$$

holds if and only if the equation

$$\text{(20)} \qquad {}^{(A)}\lim_{n \to \infty} a_n = a$$

holds.

In other words: if $\mathfrak{a}$ is an invariant subring of a Boolean ring A, then the sequential topology $\tau(\mathfrak{a})$ is the restriction of the sequential topology $\tau(A)$ to the subring $\mathfrak{a}$.

PROOF. From Definition 7 and from Theorems 46 and 49 follows immediately that (20) implies (19). Let us suppose now that (19) is satisfied. Let further b be an interelement of the sequence a_n with respect to A, and $\mathfrak{U}$ the class of those subelements of this sequence which are contained in $\mathfrak{a}$. Then, by Theorem 19,

$$\sum_{u \in \mathfrak{U}}^{(\mathfrak{a})} u = a;$$

hence, we have, by Definition 9, also

$$\sum_{u \in \mathfrak{U}}^{(A)} u = a$$

and, by Theorem 8,

$$b > a.$$

If the sequence a_n is bounded with respect to $\mathfrak{a}$, we can prove in a similar way that also $b < a$; thus we know that (19) implies (20) if the sequence a_n is bounded

with respect to $\mathfrak{a}$. If this condition is not satisfied, let c be an arbitrary element of $\mathfrak{a}$; we have, by Lemma 5,

$$^{(\mathfrak{a})}\lim_{n\to\infty} a_n c = ac$$

and, by what we have just proved,

$$^{(A)}\lim_{n\to\infty} a_n c = ac.$$

Hence we have

$$bc = ac$$

and

$$(b + ab)c = 0$$

for every element c of $\mathfrak{a}$. Thus we have especially

$$(b + ab)a_n = 0$$

for $n = 1, 2, 3, \cdots$. The last equation involves, by Theorem 5, the equation

$$(b + ab)b = 0$$

and consequently

$$b < a.$$

This proves the theorem also for the case that the sequence a_n is not bounded with respect to $\mathfrak{a}$.

From Theorem 50 follows that in the example given at the beginning of this paragraph the subring $\mathfrak{a}$ is not invariant in A.

THEOREM 51. *If $\mathfrak{a}$ is an invariant subring of a Boolean ring A, and a_n ($n = 1, 2, 3, \cdots$) and a are elements of $\mathfrak{a}$, then the equation*

$$^{(\mathfrak{a})}\underline{\lim}_{n\to\infty}\, a_n$$

holds if and only if the equation

$$^{(A)}\underline{\lim}_{n\to\infty}\, a_n$$

holds.

PROOF. In face of Theorems 44 and 50 we have only to prove that the equation

$$\underline{\lim}_{n\to\infty}\, (a_n \vee a) = a \tag{14}$$

holds with respect to $\mathfrak{a}$ if and only if it holds with respect to A. But this assertion is equivalent with the assertion that a is an interelement of the sequence

$a_n \vee a$ with respect to $\mathfrak{a}$ if and only if it is with respect to A; and that the latter assertion is true follows from Theorems 46 and 49.

THEOREM 52. *If $\mathfrak{a}$ is an invariant subring of a Boolean ring A, and a_n ($n = 1, 2, 3, \cdots$) and a are elements of $\mathfrak{a}$, then the equation*

$$^{(\mathfrak{a})}\overline{\lim_{n\to\infty}}\, a_n = a$$

holds if and only if the equation

$$^{(A)}\overline{\lim_{n\to\infty}}\, a_n = a$$

holds.

PROOF. Compare the proof of Theorem 51 and see Theorems 45, 50, 46, and 49.

THEOREM 53. *If $\mathfrak{a}_2$ is an invariant subring of a Boolean ring A, and $\mathfrak{a}_1$ an invariant subring of $\mathfrak{a}_2$, then $\mathfrak{a}_1$ is also an invariant subring of A.*

THEOREM 54. *Let $\mathfrak{B}$ be a chain of subrings of a Boolean ring A such that $\mathfrak{a} \in \mathfrak{B}$, $\mathfrak{b} \in \mathfrak{B}$, and $\mathfrak{a} \subset \mathfrak{b}$ imply that $\mathfrak{a}$ is invariant in $\mathfrak{b}$. Then each element $\mathfrak{a}$ of $\mathfrak{B}$ is also invariant in $\sum^{(\mathfrak{R})}_{\mathfrak{b} \in \mathfrak{B}} \mathfrak{b}$.* (That $\mathfrak{B}$ is a chain means that $\mathfrak{a} \in \mathfrak{B}$, $\mathfrak{b} \in \mathfrak{B}$ imply $\mathfrak{a} \subset \mathfrak{b}$ or $\mathfrak{b} \subset \mathfrak{a}$. $\mathfrak{R}$ is the lattice of all subrings of A.)

PROOF. Let $\mathfrak{a}$ be an element of $\mathfrak{B}$, $\mathfrak{C}$ a subset of $\mathfrak{a}$, and b an element of $\mathfrak{a}$. Suppose that

$$\sum_{a \in \mathfrak{C}}^{(\mathfrak{a})} a = b.$$

If c is an element of $\sum^{(\mathfrak{R})}_{\mathfrak{b} \in \mathfrak{B}} \mathfrak{b}$, and $c > a$ whenever $a \in \mathfrak{C}$, then c is contained in some element $\mathfrak{b}$ of $\mathfrak{B}$, and it can be supposed that $\mathfrak{b} \supset \mathfrak{a}$ so that $\mathfrak{a}$ is invariant in $\mathfrak{b}$. (Notice that $\sum^{(\mathfrak{R})}_{\mathfrak{b} \in \mathfrak{B}} \mathfrak{b}$ is the *set-theoretical* union of the subrings $\mathfrak{b}$ which are elements of $\mathfrak{B}$.) Hence $c > b$. This proves the theorem.

THEOREM 55. *If $\mathfrak{a}_1$ is an invariant subring of a Boolean ring A and $\mathfrak{a}_2$ a subring of A containing $\mathfrak{a}_1$, then $\mathfrak{a}_1$ is also invariant in $\mathfrak{a}_2$.*

PROOF. If $\mathfrak{B}$ is a subclass of $\mathfrak{a}_1$, b an element of $\mathfrak{a}_1$, and

$$\sum_{a \in \mathfrak{B}}^{(\mathfrak{a}_1)} a = b,$$

we have

$$\sum_{a \in \mathfrak{B}}^{(A)} a = b$$

because $\mathfrak{a}_1$ is invariant in A, and

$$\sum_{a \in \mathfrak{B}}^{(\mathfrak{a}_2)} a = b$$

by Theorem 47.

THEOREM 56. *Every ideal $\mathfrak{a}$ of a Boolean ring A is invariant in A.*

PROOF. Let $\mathfrak{B}$ be a subclass of $\mathfrak{a}$, b an element of $\mathfrak{a}$, and

$$\sum_{a \in \mathfrak{B}}^{(\mathfrak{a})} a = b.$$

If c is an arbitrary element of A, and $c > a$ whenever $a \in \mathfrak{B}$, we have also

$$bc > a$$

for every element a of $\mathfrak{B}$. Since bc is an element of $\mathfrak{a}$, we have

$$bc > b,$$

or equivalently

$$c > b$$

proving the theorem.

THEOREM 57. *If $\mathfrak{a}$ is an invariant subring of a Boolean ring A, and $\mathfrak{b}$ a normal ideal in A, then $\mathfrak{ab}$ is a normal ideal in $\mathfrak{a}$, and every normal ideal in $\mathfrak{a}$ can be obtained in this way.*

PROOF. If $\mathfrak{b}$ is a normal ideal in A, it is obvious that $\mathfrak{ab}$ is an ideal in $\mathfrak{a}$. If a is an element of the second orthocomplement of $\mathfrak{ab}$ with respect to $\mathfrak{a}$, we have, by Theorem 18,

$$a = \sum_{x \in \mathfrak{B}}^{(\mathfrak{a})} x$$

for an appropriate subclass $\mathfrak{B}$ of $\mathfrak{ab}$. Since $\mathfrak{a}$ is invariant in A, we have also

$$a = \sum_{x \in \mathfrak{B}}^{(A)} x.$$

Hence a is, again by Theorem 18, an element of $\mathfrak{b}$; since it is also an element of $\mathfrak{a}$, it is an element of $\mathfrak{ab}$. This proves the first part of the theorem.

If on the other side $\mathfrak{c}$ is an arbitrary normal ideal in $\mathfrak{a}$, we have obviously

$$\mathfrak{c} = \mathfrak{a}\mathfrak{c}''$$

the dashing referred to A, and $\mathfrak{c}''$ is a normal ideal in A.

5. Join-extensions

The relation between a partially ordered set and its completion by cuts (cf. H. MacNeille, "Partially ordered sets," Trans. Am. Math. Soc. 42 (1937), p. 445, line 9 from bottom) is a special case of the following definition.

DEFINITION 10. *If A is a subring of a Boolean ring R, then R is said to be a join-extension of A if to every element b of R there exists a subclass $\mathfrak{B}$ of A such that*

$$b = \sum_{a \in \mathfrak{B}}^{(R)} a.$$

If R is a join-extension of a Boolean ring A and a subring of a Boolean ring B, we shall say that R is a join-extension of A in B.

THEOREM 58. *If R is a join-extension of a Boolean ring A, then every element of R is the join of those elements of A which it contains, with respect to R.*

THEOREM 59. *If A and A^* are Boolean rings between which there exists an isomorphism Λ, R a join-extension of A, and R^* a join-extension of A^*, then there exists at the utmost one extension of the isomorphism Λ to an isomorphism Λ_R between R and R^*. If Λ_R exists it can be defined in the following way: let b be an element of R, $\mathfrak{a}$ the class of those elements of A which are contained in b, and $\mathfrak{a}^*$ the subclass of A^* corresponding to $\mathfrak{a}$ by virtue of Λ; then the element b^* of R^* corresponding to b by virtue of Λ_R is defined by the equation*

$$b^* = \sum_{a^* \in \mathfrak{a}^*}^{(R^*)} a^*.$$

The proof of Theorems 58 and 59 can be left to the reader.

THEOREM 60. (Corollary of Theorem 59.) *If R is a join-extension of a Boolean ring A, the identity is the only automorphism of R which carries every element of A into itself.*

THEOREM 61. *Any join-extension of a Boolean ring A is an invariant extension of A.*

PROOF. Let R be a join-extension of A, b an element of A, and $\mathfrak{B}$ a subclass of A such that

$$\sum_{a \in \mathfrak{B}}^{(A)} a = b. \tag{21}$$

Further let c be an element of R such that

$$c > a$$

for $a \in \mathfrak{B}$. Then we have also

$$(b + bc)a = 0 \tag{22}$$

for $a \in \mathfrak{B}$. On the other side $b + bc$ is itself an element of R. Hence we have, by Theorem 58,

$$b + bc = \sum_{\substack{x < b+bc \\ x \in A}}^{(R)} x. \tag{23}$$

From (22) and (23) follows

$$xa = 0 \tag{24}$$

for $x < b + bc$, $x \in A$, $a \in \mathfrak{B}$, and from (21) and (24), by Theorem 13,

$$xb = 0$$

for $x < b + bc$, $x \in A$. Now we get from (23)

$$b + bc = 0,$$

thus $c > b$. This proves

$$\sum_{a \in \mathfrak{B}}^{(R)} a = b.$$

THEOREM 62. *If R_1 is a join-extension of A, and R_2 a join-extension of R_1, then R_2 is a join-extension of A.*

PROOF. If b is an element of R_2 we have, by hypothesis and by Theorem 58,

$$b = \sum_{\substack{a<b \\ a \in R_1}}^{(R_2)} a$$

and

$$a = \sum_{\substack{x<a \\ x \in A}}^{(R_1)} x$$

for $a < b$, $a \in R_1$. Since R_1 is, by Theorem 61, invariant in R_2 we have also

$$a = \sum_{\substack{x<a \\ x \in A}}^{(R_2)} x \quad \text{for} \quad a < b, a \in R_1.$$

Hence

$$b = \sum_{\substack{a<b \\ a \in R_1}}^{(R_2)} \sum_{\substack{x<a \\ x \in A}}^{(R_2)} x.$$

This proves the theorem, according to Definition 10.

THEOREM 63. (Compare Theorem 54.) *Let $\mathfrak{B}$ be a chain of subrings of a Boolean ring B such that $R_1 \in \mathfrak{B}$, $R_2 \in \mathfrak{B}$ and $R_1 \subset R_2$ imply that R_2 is a join-extension of R_1. Then $\sum_{R \in \mathfrak{B}}^{(\mathfrak{R})} R$ is a join-extension of each element R of $\mathfrak{B}$, by $\mathfrak{R}$ denoted the lattice of all subrings of B.*

PROOF. If R_0 is an arbitrary element of $\mathfrak{B}$, and b an element of $\sum_{R \in \mathfrak{B}}^{(\mathfrak{R})} R$, then b is also contained in at least one subring R of B which is an element of $\mathfrak{B}$, and it can be supposed that $R_0 \subset R$. Hence we have

$$b = \sum_{\substack{a<b \\ a \in R_0}}^{(R)} a.$$

From this equation follows, by Theorem 54,

$$b = \sum_{\substack{a<b \\ a \in R_0}}^{(L)} a$$

if $\sum_{R \in \mathfrak{B}}^{(\mathfrak{R})} R$ is denoted by L. Hence $\sum_{R \in \mathfrak{B}}^{(\mathfrak{R})} R$ is a join-extension of R_0, proving the theorem.

THEOREM 64. *If R_2 is a join-extension of a Boolean ring A, and R_1 a subring of R_2 containing A, then R_1 is a join-extension of A, and R_2 is a join-extension of R_1.*

PROOF. If b is an element of R_1, it is also an element of R_2. Hence we have, by Theorem 58 and by hypothesis,

$$b = \sum_{\substack{a<b \\ a \in A}}^{(R_2)} a$$

and, by Theorem 47,

$$b = \sum_{\substack{a<b \\ a \in A}}^{(R_1)} a.$$

This proves that R_1 is a join-extension of A.

The second part of Theorem 64 is trivial.

If A and R are subrings of a Boolean ring B, A is invariant in B, and R a join-extension of A, then R is not necessarily invariant in B. Let B be the Boolean ring of all classes of natural numbers, A the subring of B consisting of all finite classes of natural numbers different from 1, and R the subring of B consisting of the elements of A and their complements with respect to the unit-element of B. Then A is invariant in B (it is an ideal of B), R is a join-extension of A, but R is not invariant in B.

THEOREM 65. *If A, R_1, and R_2 are subrings of a Boolean ring R_3, $A \subset R_1 \subset R_2$, and R_3 is a join-extension of A, then R_2 is a join-extension of R_1.*

PROOF. See Theorem 64.

THEOREM 66. *If A is an invariant subring of a Boolean ring B, and N the class of all elements b of B such that*

$$b = \sum_{x \in \mathfrak{B}}^{(B)} x$$

for an appropriate subclass $\mathfrak{B}$ of A (so that especially all elements of A belong to N) then N is itself a subring of B; hence N is a join-extension of A in B.

PROOF. It is obvious that $b_1 \in N$ and $b_2 \in N$ imply $b_1 b_2 \in N$ and $b_1 \vee b_2 \in N$. (See Theorem 13.)

Next we shall prove that $a \in A$ and $b \in N$ imply $a + ab \in N$. We put

$$\mathfrak{a}(a)\mathfrak{a}(b)A = \mathfrak{a},$$

the signs $\mathfrak{a}(a)$ and $\mathfrak{a}(b)$ referred to the Boolean ring B. $\mathfrak{a}(a)A$ is an ideal in A and consequently, by Theorem 56, invariant in A. Since A is, by hypothesis, invariant in B, $\mathfrak{a}(a)A$ is, by Theorem 53, as well invariant in B. Hence $\mathfrak{a}(a)\mathfrak{a}(b)A$ is, by Theorem 57, a normal ideal in $\mathfrak{a}(a)A$. Now we have

$$(\mathfrak{a} \vee_{\mathfrak{J}} \mathfrak{a}')'' = \mathfrak{a}(a)A$$

if the dashes are referred to $\mathfrak{a}(a)A$, and $\mathfrak{J}$ is the distributive lattice of the ideals in $\mathfrak{a}(a)A$. Consequently we have, by Theorem 18,

$$\sum_{x \in \mathfrak{a} \vee_{\mathfrak{J}} \mathfrak{a}'}^{(\mathfrak{a}(a)A)} x = a$$

and

$$\sum_{\substack{x \in \mathfrak{a} \\ \text{or } x \in \mathfrak{a}'}}^{(\mathfrak{a}(a)A)} x = a$$

because every element of $\mathfrak{a} \vee_{\mathfrak{J}} \mathfrak{a}'$ can be decomposed into the join of an element of $\mathfrak{a}$ and an element of $\mathfrak{a}'$. Hence we have also

$$\sum_{\substack{x \in \mathfrak{a} \\ \text{or } x \in \mathfrak{a}'}}^{(B)} x = a. \tag{25}$$

Since ab is an element of N, we have

$$ab = \sum_{x \in \mathfrak{B}}^{(B)} x$$

for an appropriate subclass $\mathfrak{B}$ of A. From this it can be easily concluded that also

$$ab = \sum_{\substack{x \in A \\ x < ab}}^{(B)} x,$$

i.e.

$$ab = \sum_{x \in \mathfrak{a}}^{(B)} x.$$

If on the other side $y \in \mathfrak{a}'$, then $y < a$ and $yx = 0$ for $x \in \mathfrak{a}$, whence $yb = 0$, and

$$y < a + ab.$$

If d is an element of B, and $d > y$ whenever $y \in \mathfrak{a}'$ then

$$d \vee ab > x$$

for $x \in \mathfrak{a}$ as well as for $x \in \mathfrak{a}'$; hence, according to (25),

$$d \vee ab > a,$$

and

$$d > a + ab.$$

This proves

$$\sum_{y \in \mathfrak{a}'}^{(B)} y = a + ab.$$

Since $\mathfrak{a}'$ is a subclass of A, $a + ab$ is an element of N, as we wished to prove.

If a and b are arbitrary elements of N, then there exists a subclass $\mathfrak{C}$ of A such that

$$\sum_{x \in \mathfrak{C}}^{(B)} x = a. \tag{26}$$

From (26) follows, by Theorem 17,

$$\sum_{x \in \mathfrak{C}}^{(B)} (x + xb) = a + ab.$$

Since $x + xb$ is, according to what we have just proved, an element of N, whenever x is an element of A, and b an element of N, we find that also $a + ab$ is an element of N. Now it is clear that also $a + b = (a + ab) \vee (ab + b)$ is an element of N, and Theorem 66 is proved.

If A is an arbitrary subring of B then the subclass N of B, defined in the same way as in Theorem 66, is certainly a *sublattice* of B, but not necessarily a *subring* of B. If, in the example given at the beginning of §4, we put for instance

$$a = a_1$$

$$b = \sum_{n=2}^{\infty}{}^{(A)} a_n,$$

then $a + ab$ is not representable as a join of elements of $\mathfrak{a}$ with respect to A.

If it is known, conversely, that N is a subring of B, and B is a complete Boolean ring, one can easily conclude that A is invariant in B. But this statement does not remain true if we omit the supposition regarding the completeness of B.

Let A be the Boolean ring whose elements are the finite sets of even positive integers and the complements of these sets with respect to the set of all positive integers, and let B consist of all finite sets of positive integers and their complements with respect to the set of all positive integers; the finite operations in A and B are meant set-theoretically. Then A is a subring of B, and N has the same property because it is identical with A, but A is not invariant in B.

Another form of Theorem 66 is

THEOREM 67. *Let A be an invariant subring of a Boolean ring B, and $\mathfrak{A}$ a subclass of A. If in this case $\prod^{(B)}_{x\in\mathfrak{A}} x$ exists then there exists also a subclass $\mathfrak{B}$ of A such that $\sum^{(B)}_{y\in\mathfrak{B}} y$ exists and is equal to $\prod^{(B)}_{x\in\mathfrak{A}} x$; and if $\sum^{(B)}_{x\in\mathfrak{A}} x$ exists, and moreover an element a of A such that $x < a$ for $x \in \mathfrak{A}$, then there exists a subclass $\mathfrak{C}$ of A such that $\prod^{(B)}_{z\in\mathfrak{C}} z$ exists and is equal to $\sum^{(B)}_{x\in\mathfrak{A}} x$.*

PROOF. If $\prod^{(B)}_{x\in\mathfrak{A}} x$ exists, and x_0 is an element of $\mathfrak{A}$, then $\sum^{(B)}_{x\in\mathfrak{A}} (x_0 + x_0 x)$ exists too and is equal to $x_0 + \prod^{(B)}_{x\in\mathfrak{A}} x$. (Theorem 14.) Consequently we have

$$\prod_{x\in\mathfrak{A}}^{(B)} x = x_0 + \sum_{x\in\mathfrak{A}}^{(B)} (x_0 + x_0 x).$$

Hence $\prod^{(B)}_{x\in\mathfrak{A}} x$ is, by Theorem 66, an element of N.

If there exist $\sum^{(B)}_{x\in\mathfrak{A}} x$ and an element a of A such that $x < a$ for $x \in \mathfrak{A}$, then there exists, by Theorem 66, a subclass $\mathfrak{C}^*$ of A such that

$$a + \sum_{x\in\mathfrak{A}}^{(B)} x = \sum_{z\in\mathfrak{C}^*}^{(B)} z;$$

hence we have, by Theorem 15,

$$\sum_{x\in\mathfrak{A}}^{(B)} x = \prod_{z\in\mathfrak{C}^*}^{(B)} (a + z).$$

THEOREM 68.[1] *The Boolean ring $\mathfrak{N}$ of the normal ideals of a Boolean ring A is a join-extension of the Boolean ring $\mathfrak{P}$ of the principal ideals of A. Hence $\mathfrak{P}$ is invariant in $\mathfrak{N}$.*

PROOF. If $\mathfrak{b}$ is a normal ideal in A, we have obviously

$$\mathfrak{b} = \sum_{a\in\mathfrak{b}}^{(\mathfrak{N})} \mathfrak{a}(a)$$

or

$$\mathfrak{b} = \sum_{\substack{\mathfrak{a}\subset\mathfrak{b} \\ \mathfrak{a}\in\mathfrak{P}}}^{(\mathfrak{N})} \mathfrak{a}.$$

[1] *Editorial comment.* Theorems 68, 69, 80, 81, and some of the other theorems of this section are either partly or wholly in H. MacNeille's "Partially ordered sets," Trans. Am. Math. Soc. 42(1937), esp. Theorem 11.13 and 11.12.

THEOREM 69. (Corollary of Theorem 68.) *Any Boolean ring can be invariantly extended to a complete Boolean ring.*

THEOREM 70. *If the signs A, B, and N have the same meaning as in Theorem* 66 *then N is invariant in B.*

PROOF. Let $\bar{B}$ be a complete Boolean ring which is an invariant extension of B. (That such a Boolean ring $\bar{B}$ exists follows from Theorem 69.) Then A is, by Theorem 53, an invariant subring of $\bar{B}$. Further let $\bar{N}$ be the subring of $\bar{B}$ arising from A and $\bar{B}$ in the same way as N from A and B in Theorem 66. N is a subring of $\bar{N}$. If now $\mathfrak{B}$ is a subclass of N, and b an element of N such that

$$\sum_{a \in \mathfrak{B}}^{(N)} a = b \tag{27}$$

then we put

$$\sum_{a \in \mathfrak{B}}^{(\bar{B})} a = c. \tag{28}$$

($\sum_{a \in \mathfrak{B}}^{(\bar{B})} a$ exists because $\bar{B}$ is a complete Boolean ring.) c satisfies the inequality

$$c < b$$

and is an element of $\bar{N}$. Since b is an element of $\bar{N}$, the same holds also for $b + c$. Hence we have

$$b + c = \sum_{x \in \mathfrak{C}}^{(\bar{B})} x \tag{29}$$

for an appropriate subclass $\mathfrak{C}$ of A. On the other side we get from (28)

$$(b + c)a = 0$$

for $a \in \mathfrak{B}$. Consequently we have the more

$$xa = 0$$

for $x \in \mathfrak{C}$, $a \in \mathfrak{B}$. Because $\mathfrak{C}$ is a subclass of N we get from this and from (27)

$$xb = 0$$

for $x \in \mathfrak{C}$. Now from (29) follows

$$b + c = 0$$

i.e. $c = b$. By this is proved that N is an invariant subring of $\bar{B}$. Hence N is, by Theorem 55, also invariant in B.

THEOREM 71. *If the signs A, B, and N have the same meaning as in Theorem* 66, *and R is a join-extension of A invariant in B* (i.e. an invariant subring of B which is a join-extension of A), *then R is contained in N.*

DEFINITION 11. *If the signs A, B, and N have the same meaning as in Theorem* 66, *then N is called the maximal join-extension of A invariant in B.*

Definition 11 is justified by Theorem 71.

THEOREM 72. *If A is an invariant subring of a Boolean ring B, the join-extensions of A invariant in B form a complete sublattice of the lattice of all sub-*

rings of B. The zero of this sublattice is A, and its unit is the maximal join-extension of A invariant in B.

THEOREM 73. *Let A be an invariant subring of Boolean ring B, R a join-extension of A invariant in B, N the maximal join-extension of A invariant in B, and D an invariant subring of B containing N. In this case N is also the maximal join-extension of R invariant in D.*

THEOREM 74. *If $\mathfrak{a}$ is an ideal of a Boolean ring A, the maximal join-extension of $\mathfrak{a}$ invariant in A is identical with $\mathfrak{a}''$.*

PROOF. See Theorem 18.

If $\mathfrak{a}$ is an arbitrary invariant subring of A, the maximal join-extension of $\mathfrak{a}$ invariant in A is certainly a subring of $\mathfrak{a}''$ (and, by Theorem 55, even invariant in $\mathfrak{a}''$), but not necessarily identical with $\mathfrak{a}''$. Let for instance A be a *finite* Boolean ring, and $\mathfrak{a}$ a proper subring of A which contains the unit of A. Then $\mathfrak{a}$ is certainly invariant in A, and $\mathfrak{a}'' = A$, while the maximal join-extension of $\mathfrak{a}$ invariant in A is $\mathfrak{a}$ itself.

THEOREM 75. *If A is an invariant subring of a complete Boolean ring B, the maximal join-extension N of A invariant in B is itself a complete Boolean ring, and any complete invariant subring of B containing A contains N.*

PROOF. If $\mathfrak{B}$ is a subclass of N then $\sum^{(B)}_{x \in \mathfrak{B}} x$ exists and is, according to the definition of N, contained in N; hence $\sum^{(N)}_{x \in \mathfrak{B}} x$ exists and is identical with $\sum^{(B)}_{x \in \mathfrak{B}} x$. (See Theorem 47.) If we put

$$\sum_{x \in \mathfrak{B}}^{(N)} x = a$$

then $\sum^{(N)}_{x \in \mathfrak{B}} (a + x)$ exists too. Hence, by Theorem 15, there exists also $\prod^{(N)}_{x \in \mathfrak{B}} x$ and is identical with $a + \sum^{(N)}_{x \in \mathfrak{B}} (a + x)$.

If D is a complete invariant subring of B, and D contains A, then $\sum^{(D)}_{x \in \mathfrak{A}} x$ exists whenever $\mathfrak{A}$ is a subclass of A, and we have

$$\sum_{x \in \mathfrak{A}}^{(D)} x = \sum_{x \in \mathfrak{A}}^{(B)} x.$$

Thus D contains all elements of N.

DEFINITION 12. *An extension N of a Boolean ring A is said to be a minimal complete invariant extension of A, or shorter: a completion of A, if N is a complete Boolean ring invariant over A, and there is no complete proper subring of N containing A.*

THEOREM 76. *If A is an invariant subring of a complete Boolean ring B, and N the maximal join-extension of A invariant in B, then N is a minimal complete invariant extension of A.*

PROOF. N is, by Theorem 75, a complete Boolean ring and, by Theorems 61 and 66, invariant over A. If R is a complete subring of N containing A then R is, by Theorems 53 and 64, invariant in B; hence we have, by Theorem 75,

$$R \supset N$$

and consequently

$$R = N.$$

THEOREM 77. *A Boolean ring N is a minimal complete invariant extension of a Boolean ring A if and only if N is a complete Boolean ring and a join-extension of A.*

PROOF. If N is a minimal complete invariant extension of A, from Theorem 75 and Definition 12 follows that the maximal join-extension of A invariant in N cannot be different from N; hence N is a join-extension of A.

If N is complete and a join-extension of A, we see immediately that N is the maximal join-extension of A invariant in N; hence N is, by Theorem 76, a minimal complete invariant extension of A.

THEOREM 78. *If A and N are invariant subrings of a Boolean ring B, and N is a minimal complete invariant extension of A, then N is the maximal join-extension of A invariant in B.*

The proof follows easily from Theorem 77 and Definition 11.

THEOREM 79. *Any Boolean ring has at least one minimal complete invariant extension.*

PROOF. See Theorems 69 and 76.

THEOREM 80. *The Boolean ring $\mathfrak{N}$ of the normal ideals of a Boolean ring A is a minimal complete invariant extension of the Boolean ring $\mathfrak{P}$ of the principal ideals of A.*

PROOF. See Theorems 68 and 77.

THEOREM 81. *Any minimal complete invariant extension N of a Boolean ring A is isomorphic with the Boolean ring $\mathfrak{N}$ of the normal ideals of A, and there is exactly one isomorphism between N and $\mathfrak{N}$ carrying every element a of A into the principal ideal $\mathfrak{a}(a)$ of A.*

PROOF. Since N is a complete Boolean ring, every normal ideal of N is principal according to Theorem 18. If a is an element of N, $\mathfrak{b}(a)$ the principal ideal of N generated by a, and we put

$$\mathfrak{a} = A \cdot \mathfrak{b}(a) \tag{31}$$

then $\mathfrak{a}$ is a normal ideal of A, and any normal ideal of A can be obtained in this way. (Theorem 57.) On the other side N is, by Theorem 77, a join-extension of A. Therefore we get from (31)

$$a = \sum_{x \in \mathfrak{a}}^{(N)} x. \tag{32}$$

(See Theorem 58.) We see that to different elements a of N there correspond different normal ideals $\mathfrak{a}$ of $\mathfrak{N}$. Finally it is obvious that the biunivocal correspondence between N and $\mathfrak{N}$ defined by (31) or (32) preserves inclusion. Hence it is an isomorphism between N and $\mathfrak{N}$. If $a \in A$, then $A \cdot \mathfrak{b}(a) = \mathfrak{a}(a)$. That there is but one such isomorphism follows from Theorems 59 and 68.

THEOREM 82. *If A and A^* are Boolean rings between which there exists an isomorphism Λ, N a minimal complete invariant extension of A, and N^* a minimal complete invariant extension of A^*, then there exists exactly one extension of the isomorphism Λ to an isomorphism Λ_N between N and N^*.*

In particular we may say that any two minimal complete invariant extensions of the same Boolean ring are abstractly identical.

The proof of Theorem 82 follows easily from Theorems 59 and 81.

Now we can state the following generalization of Theorem 59.

THEOREM 83. *If A and A^* are Boolean rings between which there exists an isomorphism Λ, R a join-extension of A, and $\hat{R}$ a join-extension of A^*, then there exists at the utmost one subring R^* of $\hat{R}$ containing A^* such that the isomorphism Λ can be extended to an isomorphism Λ_R between R and R^*. R^* exists if and only if $\sum_{a^* \in \mathfrak{a}^*}^{(\hat{R})} a^*$ exists whenever b is an element of R, and $\mathfrak{a}^*$ is the subclass of A^* assigned to the element b of R in the same way as in Theorem 59. In this case R^* is the class of all such joins $\sum_{a^* \in \mathfrak{a}^*}^{(\hat{R})} a^*$, the isomorphism Λ_R is unique, and the element b^* of R^*, assigned to the element b of R by virtue of Λ_R, is given by the equation*

$$b^* = \sum_{a^* \in \mathfrak{a}^*}^{(\hat{R})} a^*.$$

Hence if $\hat{R}$ is especially a completion of A^*, then R^* certainly exists.

PROOF. In face of Theorems 59, 61, and 64 we may restrict ourselves to prove that the existence of all joins $\sum_{a^* \in \mathfrak{a}^*}^{(\hat{R})} a^*$ implies that R^* exists. Let N be a completion of R, and N^* a completion of $\hat{R}$. (See Theorem 79.) It is obvious that N and N^* are also completions of A and A^* respectively. Hence the isomorphism Λ can be extended to an isomorphism Λ_N between N and N^* in exactly one way. (Theorem 82.) And in face of Theorem 59 it is obvious that the existence of all our joins $\sum_{a^* \in \mathfrak{a}^*}^{(\hat{R})} a^*$ implies that the subring R^* of N^* into which R is carried by Λ_N is contained in $\hat{R}$. Thus we have proved what we had wished to.

In the rest of this paragraph we shall state some corollaries of Theorems 46, 49, 50, 51, 52, and 68. Here and in the next paragraph we shall consider a fixed Boolean ring A; the Boolean ring of the principal ideals of A will be denoted by $\mathfrak{P}$, and the Boolean ring of the normal ideals of A by $\mathfrak{N}$. Notice Theorem 38.

THEOREM 84. *An element a is an interelement of a sequence a_n if and only if*

$$^{(\mathfrak{N})}\underline{\lim}_{n\to\infty}\, \mathfrak{a}(a_n) \subset \mathfrak{a}(a) \subset {}^{(\mathfrak{N})}\overline{\lim}_{n\to\infty}\, \mathfrak{a}(a_n). \tag{33}$$

PROOF. See Theorems 11, 46, 49, and 68.

THEOREM 85. $\underline{\lim}_{n\to\infty} a_n$ *exists if and only if* $^{(\mathfrak{N})}\underline{\lim}_{n\to\infty} \mathfrak{a}(a_n)$ *is a principal ideal* $\mathfrak{a}(a)$; *in this case we have*

$$\underline{\lim}_{n\to\infty} a_n = a.$$

PROOF. See Theorems 51 and 68.

THEOREM 86. $\overline{\lim}_{n\to\infty} a_n$ *exists if and only if* $^{(\mathfrak{N})}\overline{\lim}_{n\to\infty} \mathfrak{a}(a_n)$ *is a principal ideal* $\mathfrak{a}(a)$; *in this case we have*

$$\overline{\lim}_{n\to\infty} a_n = a.$$

PROOF. See Theorems 52 and 68.

THEOREM 87. *A sequence of elements a_n of a Boolean ring A is convergent with*

respect to A if and only if the sequence of the principal ideals $\mathfrak{a}(a_n)$ is convergent with respect to $\mathfrak{N}$, and its limit is a principal ideal $\mathfrak{a}(a)$; in this case we have

$$^{(A)}\lim_{n\to\infty} a_n = a.$$

PROOF. See Theorems 50 and 68.

At this place we mention a theorem which may be considered as a generalization of Theorem 20:

THEOREM 88. *If $\mathfrak{U}$ is the class of the subelements of a sequence a_n, we have*

$$^{(\mathfrak{N})}\underline{\lim}_{n\to\infty}\, \mathfrak{a}(a_n) = \mathfrak{U}''. \tag{34}$$

PROOF. It is evident that

$$\sum_{n=1}^{\infty}{}^{(\mathfrak{J})} \prod_{k=n}^{\infty}{}^{(\mathfrak{N})} \mathfrak{a}(a_k) = \mathfrak{U}$$

if $\mathfrak{J}$ is the distributive lattice of the ideals of the given Boolean ring, and from this and from Theorem 36 there ensues immediately the asserted Relation (34).

6. Relations between Interelements, Lower and Upper Limits, and Limits

THEOREM 89. *If $a_n < b_n$ for $n = 1, 2, 3, \cdots$, a is an interelement of the sequence a_n, and $\overline{\lim}_{n\to\infty} b_n$ exists, then*

$$a < \overline{\lim}_{n\to\infty} b_n. \tag{35}$$

THEOREM 90. *If $a_n < b_n$ for $n = 1, 2, 3, \cdots$, $\underline{\lim}_{n\to\infty} a_n$ exists, and b is an interelement of the sequence b_n, then*

$$\underline{\lim}_{n\to\infty} a_n < b. \tag{36}$$

PROOF OF THEOREMS 89 AND 90. It is obvious that $a_n < b_n$ for $n = 1, 2, 3, \cdots$ implies

$$^{(\mathfrak{N})}\underline{\lim}_{n\to\infty}\, \mathfrak{a}(a_n) \subset {}^{(\mathfrak{N})}\underline{\lim}_{n\to\infty}\, \mathfrak{a}(b_n)$$

and

$$^{(\mathfrak{N})}\overline{\lim}_{n\to\infty}\, \mathfrak{a}(a_n) \subset {}^{(\mathfrak{N})}\overline{\lim}_{n\to\infty}\, \mathfrak{a}(b_n).$$

After this is settled Theorems 89 and 90 can be easily concluded from Theorem 84.

THEOREM 91. *If $a_n < b_n$ for $n = 1, 2, 3, \cdots$ then those of the relations*

$$\underline{\lim}_{n\to\infty} a_n < \underline{\lim}_{n\to\infty} b_n$$

$$\underline{\lim}_{n\to\infty} a_n < \overline{\lim}_{n\to\infty} b_n$$

$$\overline{\lim}_{n\to\infty} a_n < \overline{\lim}_{n\to\infty} b_n$$

hold in which both sides exist.

PROOF. See Theorems 89 and 90.

THEOREM 92. *If* $\lim_{n\to\infty} b_n = 0$, *and* $a_n < b_n$ *for* $n = 1, 2, 3, \cdots$ *then also*

$$\lim_{n\to\infty} a_n = 0.$$

Notice that 0 is the only subelement and the only interelement of the sequence a_n.

LEMMA 8. *If* $a_k < a_l$ *and* $b_k < b_l$ *for* $k < l$, $k, l = 1, 2, 3, \cdots$ *and* $\sum_{n=1}^{\infty} a_n$ *and* $\sum_{n=1}^{\infty} b_n$ *exist, then* $\sum_{n=1}^{\infty} a_n b_n$ *exists too and is equal to* $\sum_{n=1}^{\infty} a_n \cdot \sum_{n=1}^{\infty} b_n$.

PROOF. We have obviously

$$\sum_{n=1}^{\infty} a_n \cdot \sum_{n=1}^{\infty} b_n > a_k b_k$$

for every k. If $c > a_n b_n$ for every n then

$$c > a_k b_l$$

for $k, l = 1, 2, 3, \cdots$. From this inequality we get by using twice Theorem 13 the inequality

$$c > \sum_{n=1}^{\infty} a_n \cdot \sum_{n=1}^{\infty} b_n$$

which proves the assertion.

LEMMA 9. *If* $a_k > a_l$ *and* $b_k > b_l$ *for* $k < l$, $k, l = 1, 2, 3, \cdots$ *and* $\prod_{n=1}^{\infty} a_n$ *and* $\prod_{n=1}^{\infty} b_n$ *exist, then* $\prod_{n=1}^{\infty} (a_n \vee b_n)$ *exists too and is equal to* $\prod_{n=1}^{\infty} a_n \vee \prod_{n=1}^{\infty} b_n$.

PROOF. $\prod_{n=1}^{\infty} a_n \vee \prod_{n=1}^{\infty} b_n < a_k \vee b_k$ for every k. If $c < a_n \vee b_n$ for every n then $c < a_k \vee b_l$ for $k, l = 1, 2, 3, \cdots$. Hence we get by using twice Theorem 12

$$c < \prod_{n=1}^{\infty} a_n \vee \prod_{n=1}^{\infty} b_n.$$

This proves the assertion.

THEOREM 93. *If* $\underline{\lim}_{n\to\infty} a_n$ *and* $\underline{\lim}_{n\to\infty} b_n$ *exist then* $\underline{\lim}_{n\to\infty} a_n b_n$ *exists too and is equal to* $\underline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n$.

PROOF.

$$\begin{aligned} {}^{(\mathfrak{R})}\underline{\lim}_{n\to\infty} \mathfrak{a}(a_n b_n) &= {}^{(\mathfrak{R})}\underline{\lim}_{n\to\infty} \mathfrak{a}(a_n)\mathfrak{a}(b_n) = \sum_{n=1}^{\infty}{}^{(\mathfrak{R})} \prod_{k=n}^{\infty}{}^{(\mathfrak{R})} \mathfrak{a}(a_k)\mathfrak{a}(b_k) \\ &= \sum_{n=1}^{\infty}{}^{(\mathfrak{R})} \left[\prod_{k=n}^{\infty}{}^{(\mathfrak{R})} \mathfrak{a}(a_k) \cdot \prod_{k=n}^{\infty} \mathfrak{a}(b_k) \right]; \end{aligned}$$

hence, by Lemma 8,

$${}^{(\mathfrak{R})}\underline{\lim}_{n\to\infty} \mathfrak{a}(a_n b_n) = \sum_{n=1}^{\infty}{}^{(\mathfrak{R})} \prod_{k=n}^{\infty}{}^{(\mathfrak{R})} \mathfrak{a}(a_k) \cdot \sum_{n=1}^{\infty}{}^{(\mathfrak{R})} \prod_{k=n}^{\infty}{}^{(\mathfrak{R})} \mathfrak{a}(b_k) = {}^{(\mathfrak{R})}\underline{\lim}_{n\to\infty} \mathfrak{a}(a_n) \cdot {}^{(\mathfrak{R})}\underline{\lim}_{n\to\infty} \mathfrak{a}(b_n),$$

and, by Theorem 85,

$$^{(\mathfrak{R})}\varliminf_{n\to\infty} \mathfrak{a}(a_n b_n) = \mathfrak{a}(\varliminf_{n\to\infty} a_n)\cdot\mathfrak{a}(\varliminf_{n\to\infty} b_n) = \mathfrak{a}(\varliminf_{n\to\infty} a_n\cdot\varliminf_{n\to\infty} b_n);$$

from this follows, again by Theorem 85, the assertion.

THEOREM 94. *If* $\varlimsup_{n\to\infty} a_n$ *and* $\varlimsup_{n\to\infty} b_n$ *exist then* $\varlimsup_{n\to\infty} (a_n \vee b_n)$ *exists too and is equal to* $\varlimsup_{n\to\infty} a_n \vee \varlimsup_{n\to\infty} b_n$.

PROOF. Compare the proof of Theorem 93 and see Lemma 9 and Theorem 86.

THEOREM 95. *If* $\varliminf_{n\to\infty} a_n$ *and* $\varlimsup_{n\to\infty} b_n$ *exist then* $\varliminf_{n\to\infty} (a_n + a_n b_n)$ *exists too and is equal to* $\varliminf_{n\to\infty} a_n + \varliminf_{n\to\infty} a_n\cdot\varlimsup_{n\to\infty} b_n$.

PROOF. We have

$$\mathfrak{a}(a_n + a_n b_n) = \mathfrak{a}(a_n)\cdot\mathfrak{a}'(b_n)$$

and, by Lemma 6,

$$^{(\mathfrak{R})}\varliminf_{n\to\infty} \mathfrak{a}'(b_n) = [^{(\mathfrak{R})}\varlimsup_{n\to\infty} \mathfrak{a}(b_n)]';$$

hence we have, by Theorem 93,

$$^{(\mathfrak{R})}\varliminf_{n\to\infty} \mathfrak{a}(a_n + a_n b_n) = \mathfrak{a}(\varliminf_{n\to\infty} a_n)\cdot\mathfrak{a}'(\varlimsup_{n\to\infty} b_n) = \mathfrak{a}(\varliminf_{n\to\infty} a_n + \varliminf_{n\to\infty} a_n\cdot\varlimsup_{n\to\infty} b_n).$$

THEOREM 96. *If a is an interelement of the sequence a_n, and b an interelement of the sequence b_n, then those of the inequalities*

$$ab > \varliminf_{n\to\infty} a_n b_n,$$

$$a \vee b < \varlimsup_{n\to\infty} (a_n \vee b_n),$$

and

$$a + ab > \varliminf_{n\to\infty} (a_n + a_n b_n)$$

hold the right-hand sides of which exist.

PROOF. See Theorems 84, 93, 94, and 95.

THEOREM 97. *If a is an interelement of the sequence a_n, and $\varliminf_{n\to\infty} b_n$ exists, then $a \varliminf_{n\to\infty} b_n$ is an interelement of the sequence $a_n b_n$.*

PROOF. We have for $l = 1, 2, 3, \cdots$

$$^{(\mathfrak{R})}\varlimsup_{n\to\infty} \mathfrak{a}(a_n b_n) = {}^{(\mathfrak{R})}\varlimsup_{n\to\infty} [\mathfrak{a}(a_n)\mathfrak{a}(b_n)] = \prod_{n=1}^{\infty}{}^{(\mathfrak{R})} \sum_{k=n}^{\infty}{}^{(\mathfrak{R})} \mathfrak{a}(a_k)\mathfrak{a}(b_k)$$

$$= \prod_{n=l}^{\infty}{}^{(\mathfrak{R})} \sum_{k=n}^{\infty}{}^{(\mathfrak{R})} \mathfrak{a}(a_k)\mathfrak{a}(b_k) \supset \left[\prod_{n=l}^{\infty}{}^{(\mathfrak{R})} \sum_{k=n}^{\infty}{}^{(\mathfrak{R})} \mathfrak{a}(a_k)\right]\cdot\prod_{m=l}^{\infty}{}^{(\mathfrak{R})} \mathfrak{a}(b_m)$$

$$= {}^{(\mathfrak{R})}\varlimsup_{n\to\infty} \mathfrak{a}(a_n)\cdot\prod_{m=l}^{\infty}{}^{(\mathfrak{R})} \mathfrak{a}(b_m),$$

hence

$$^{(\mathfrak{N})}\overline{\lim_{n\to\infty}}\, \mathfrak{a}(a_n b_n) \supset {}^{(\mathfrak{N})}\overline{\lim_{n\to\infty}}\, \mathfrak{a}(a_n) \cdot {}^{(\mathfrak{N})}\underline{\lim_{n\to\infty}}\, \mathfrak{a}(b_n) = {}^{(\mathfrak{N})}\overline{\lim_{n\to\infty}}\, \mathfrak{a}(a_n) \cdot \mathfrak{a}(\underline{\lim_{n\to\infty}}\, b_n)$$

$$\supset \mathfrak{a}(a) \cdot (\underline{\lim_{n\to\infty}}\, b_n) = \mathfrak{a}(a \underline{\lim_{n\to\infty}}\, b_n).$$

On the other side it follows from the first relation of Theorem 96 that

$$\mathfrak{a}(a) \cdot \mathfrak{a}(\underline{\lim_{n\to\infty}}\, b_n) \supset {}^{(\mathfrak{N})}\underline{\lim_{n\to\infty}}\, [\mathfrak{a}(a_n)\mathfrak{a}(b_n)],$$

or

$$\mathfrak{a}(a \underline{\lim_{n\to\infty}}\, b_n) \supset {}^{(\mathfrak{N})}\underline{\lim_{n\to\infty}}\, \mathfrak{a}(a_n b_n).$$

This proves the assertion, according to Theorem 84.

THEOREM 98. (Corollary of Theorem 97.) *If* $\overline{\lim_{n\to\infty}}\, a_n$ *and* $\underline{\lim_{n\to\infty}}\, b_n$ *exist then* $\overline{\lim_{n\to\infty}}\, a_n \cdot \underline{\lim_{n\to\infty}}\, b_n$ *is an interelement of the sequence* $a_n b_n$.

THEOREM 99. *If* a *is an interelement of the sequence* a_n, *and* $\overline{\lim_{n\to\infty}}\, b_n$ *exists, then* $a \vee \overline{\lim_{n\to\infty}}\, b_n$ *is an interelement of the sequence* $a_n \vee b_n$.

Theorem 99 can be proved in a way dually corresponding to the proof of Theorem 97.

THEOREM 100. (Corollary of Theorem 99.) *If* $\underline{\lim_{n\to\infty}}\, a_n$ *and* $\overline{\lim_{n\to\infty}}\, b_n$ *exist then* $\underline{\lim_{n\to\infty}}\, a_n \vee \overline{\lim_{n\to\infty}}\, b_n$ *is an interelement of the sequence* $a_n \vee b_n$.

THEOREM 101. *If* a *is an interelement of the sequence* a_n, *and* $\overline{\lim_{n\to\infty}}\, b_n$ *exists, then* $a + a \overline{\lim_{n\to\infty}}\, b_n$ *is an interelement of the sequence* $a_n + a_n b_n$.

PROOF. We have

$$\mathfrak{a}(a + a \overline{\lim_{n\to\infty}}\, b_n) = \mathfrak{a}(a) \cdot \overline{\lim_{n\to\infty}}\, \mathfrak{a}'(b_n)$$

(see Lemma 6); hence $\mathfrak{a}(a + a \overline{\lim_{n\to\infty}}\, b_n)$ is, by Theorem 97, an interelement of the sequence $\mathfrak{a}(a_n)\mathfrak{a}'(b_n) = \mathfrak{a}(a_n + a_n b_n)$ with respect to $\mathfrak{N}$.

THEOREM 102. (Corollary of Theorem 101.) *If* $\overline{\lim}\, a_n$ *and* $\overline{\lim}\, b_n$ *exist then* $\overline{\lim_{n\to\infty}}\, a_n + \overline{\lim_{n\to\infty}}\, a_n \cdot \overline{\lim_{n\to\infty}}\, b_n$ *is an interelement of the sequence* $a_n + a_n b_n$.

THEOREM 103. *If* $\underline{\lim_{n\to\infty}}\, a_n$ *exists, and* b *is an interelement of the sequence* b_n, *then* $\underline{\lim_{n\to\infty}}\, a_n + b \underline{\lim_{n\to\infty}}\, a_n$ *is an interelement of the sequence* $a_n + a_n b_n$.

The proof follows from the equation

$$\mathfrak{a}(\underline{\lim_{n\to\infty}}\, a_n + b \underline{\lim_{n\to\infty}}\, a_n) = {}^{(\mathfrak{N})}\underline{\lim_{n\to\infty}}\, \mathfrak{a}(a_n) \cdot \mathfrak{a}'(b)$$

and from Theorem 97; for $\mathfrak{a}'(b)$ is, by Lemma 7, an interelement of the sequence $\mathfrak{a}'(b_n)$ with respect to $\mathfrak{N}$.

THEOREM 104. (Corollary of Theorem 103.) *If* $\underline{\lim}_{n\to\infty} a_n$ *and* $\underline{\lim}_{n\to\infty} b_n$ *exist then* $\underline{\lim}_{n\to\infty} a_n + \underline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n$ *is an interelement of the sequence* $a_n + a_n b_n$.

THEOREM 105. *If* $\overline{\lim}_{n\to\infty} a_n$ *and* $\overline{\lim}_{n\to\infty} b_n$ *exist, and* x *is an interelement of the sequence* $a_n b_n$, *then*

$$x < \overline{\lim_{n\to\infty}} a_n \cdot \overline{\lim_{n\to\infty}} b_n .$$

PROOF. We have, by Theorem 89,

$$x < \overline{\lim_{n\to\infty}} a_n$$

and

$$x < \overline{\lim_{n\to\infty}} b_n ;$$

hence

$$x < \overline{\lim_{n\to\infty}} a_n \cdot \overline{\lim_{n\to\infty}} b_n .$$

THEOREM 106. *If* $\underline{\lim}_{n\to\infty} a_n$ *and* $\underline{\lim}_{n\to\infty} b_n$ *exist, and* x *is an interelement of the sequence* $a_n \vee b_n$, *then*

$$x > \underline{\lim_{n\to\infty}} a_n \vee \underline{\lim_{n\to\infty}} b_n .$$

PROOF. We have, by Theorem 90,

$$x > \underline{\lim_{n\to\infty}} a_n$$

and

$$x > \underline{\lim_{n\to\infty}} b_n ;$$

hence

$$x > \underline{\lim_{n\to\infty}} a_n \vee \underline{\lim_{n\to\infty}} b_n .$$

THEOREM 107. *If* $\overline{\lim}_{n\to\infty} a_n$ *and* $\underline{\lim}_{n\to\infty} b_n$ *exist, and* x *is an interelement of the sequence* $a_n + a_n b_n$, *then*

$$x < \overline{\lim_{n\to\infty}} a_n + \overline{\lim_{n\to\infty}} a_n \cdot \underline{\lim_{n\to\infty}} b_n .$$

PROOF. Since $\mathfrak{a}(x)$ is an interelement of the sequence $\mathfrak{a}(a_n + a_n b_n) = \mathfrak{a}(a_n)\mathfrak{a}'(b_n)$ with respect to $\mathfrak{R}$, we have, by Theorem 105 and Lemma 6,

$$\mathfrak{a}(x) \subset {}^{(\mathfrak{R})}\overline{\lim_{n\to\infty}}\, \mathfrak{a}(a_n) \cdot {}^{(\mathfrak{R})}\overline{\lim_{n\to\infty}}\, \mathfrak{a}'(b_n) = {}^{(\mathfrak{R})}\overline{\lim_{n\to\infty}}\, \mathfrak{a}(a_n) \cdot \left({}^{(\mathfrak{R})}\underline{\lim_{n\to\infty}}\, \mathfrak{a}(b_n)\right)'$$
$$= \mathfrak{a}\left(\overline{\lim_{n\to\infty}} a_n\right) \cdot \mathfrak{a}'\left(\underline{\lim_{n\to\infty}} b_n\right) = \mathfrak{a}\left(\overline{\lim_{n\to\infty}} a_n + \overline{\lim_{n\to\infty}} a_n \cdot \underline{\lim_{n\to\infty}} b_n\right).$$

Theorem 108. *Those of the following inclusion-relations* (37) *till* (53) *which have a sense are valid:*

$$\underline{\lim}_{n\to\infty} a_n b_n < \overline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n \tag{37}$$

$$\underline{\lim}_{n\to\infty} a_n b_n < \overline{\lim}_{n\to\infty} a_n \cdot \overline{\lim}_{n\to\infty} b_n \tag{38}$$

$$\overline{\lim}_{n\to\infty} a_n b_n > \overline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n \tag{39}$$

$$\overline{\lim}_{n\to\infty} a_n b_n < \overline{\lim}_{n\to\infty} a_n \cdot \overline{\lim}_{n\to\infty} b_n \tag{40}$$

$$\overline{\lim}_{n\to\infty} (a_n \vee b_n) > \underline{\lim}_{n\to\infty} a_n \vee \overline{\lim}_{n\to\infty} b_n \tag{41}$$

$$\overline{\lim}_{n\to\infty} (a_n \vee b_n) > \underline{\lim}_{n\to\infty} a_n \vee \underline{\lim}_{n\to\infty} b_n \tag{42}$$

$$\underline{\lim}_{n\to\infty} (a_n \vee b_n) < \underline{\lim}_{n\to\infty} a_n \vee \overline{\lim}_{n\to\infty} b_n \tag{43}$$

$$\underline{\lim}_{n\to\infty} (a_n \vee b_n) > \underline{\lim}_{n\to\infty} a_n \vee \underline{\lim}_{n\to\infty} b_n \tag{44}$$

$$\underline{\lim}_{n\to\infty} (a_n + a_n b_n) < \underline{\lim}_{n\to\infty} a_n + \underline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n \tag{45}$$

$$\underline{\lim}_{n\to\infty} (a_n + a_n b_n) < \overline{\lim}_{n\to\infty} a_n + \overline{\lim}_{n\to\infty} a_n \cdot \overline{\lim}_{n\to\infty} b_n \tag{46}$$

$$\underline{\lim}_{n\to\infty} (a_n + a_n b_n) < \overline{\lim}_{n\to\infty} a_n + \overline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n \tag{47}$$

$$\overline{\lim}_{n\to\infty} (a_n + a_n b_n) > \underline{\lim}_{n\to\infty} a_n + \underline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n \tag{48}$$

$$\lim_{n\to\infty} (a_n + a_n b_n) > \overline{\lim}_{n\to\infty} a_n + \overline{\lim}_{n\to\infty} a_n \cdot \overline{\lim}_{n\to\infty} b_n \tag{49}$$

$$\overline{\lim}_{n\to\infty} (a_n + a_n b_n) < \overline{\lim}_{n\to\infty} a_n + \overline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n \tag{50}$$

$$\underline{\lim}_{n\to\infty} (a_n + b_n) < \underline{\lim}_{n\to\infty} a_n + \overline{\lim}_{n\to\infty} b_n \tag{51}$$

$$\overline{\lim}_{n\to\infty} (a_n + b_n) > \underline{\lim}_{n\to\infty} a_n + \underline{\lim}_{n\to\infty} b_n \tag{52}$$

$$\overline{\lim}_{n\to\infty} (a_n + b_n) > \overline{\lim}_{n\to\infty} a_n + \overline{\lim}_{n\to\infty} b_n \tag{53}$$

Proof.

Relation	(37)	(38)	(39)	(40)	(41)	(42)	(43)	(44)	(45)	(46)	(47)	(48)	(49)	(50)
Follows from Theorem	98	105	98	105	100	106	100	106	104	102	107	104	102	107

If $\underline{\lim}_{n\to\infty} a_n$, $\overline{\lim}_{n\to\infty} b_n$, and $\underline{\lim}_{n\to\infty} (a_n + b_n)$ exist then $\underline{\lim}_{n\to\infty} (a_n + a_nb_n)$ exists too and is equal to $\underline{\lim}_{n\to\infty} a_n + \underline{\lim}_{n\to\infty} a_n \cdot \overline{\lim}_{n\to\infty} b_n$. (Theorem 95.) On the other side we have $a_n + a_nb_n = (a_n + b_n)a_n$ whence

$$\underline{\lim}_{n\to\infty} (a_n + a_nb_n) = \underline{\lim}_{n\to\infty} (a_n + b_n) \cdot \underline{\lim}_{n\to\infty} a_n .$$

(Theorem 93.) Thus we have

$$\underline{\lim}_{n\to\infty} (a_n + b_n)\, \underline{\lim}_{n\to\infty} a_n = \underline{\lim}_{n\to\infty} a_n + \underline{\lim}_{n\to\infty} a_n \cdot \overline{\lim}_{n\to\infty} b_n . \tag{54}$$

Further $\underline{\lim}_{n\to\infty} (a_n + b_n)\, \overline{\lim}_{n\to\infty} b_n$ is, by Theorem 98, an interelement of the sequence $(a_n + b_n)b_n = a_nb_n + b_n$. Hence we have, by Theorem 107,

$$\underline{\lim}_{n\to\infty} (a_n + b_n)\, \overline{\lim}_{n\to\infty} b_n < \underline{\lim}_{n\to\infty} a_n \cdot \overline{\lim}_{n\to\infty} b_n + \overline{\lim}_{n\to\infty} b_n . \tag{55}$$

From (54) and (55) we get

$$\underline{\lim}_{n\to\infty} (a_n + b_n) \cdot (\underline{\lim}_{n\to\infty} a_n \vee \overline{\lim}_{n\to\infty} b_n) < \underline{\lim}_{n\to\infty} a_n + \overline{\lim}_{n\to\infty} b_n . \tag{56}$$

Since $\underline{\lim}_{n\to\infty} a_n \vee \overline{\lim}_{n\to\infty} b_n$ is, by Theorem 100, an interelement of the sequence $a_n \vee b_n$, and $a_n + b_n < a_n \vee b_n$, we have, by Theorem 90,

$$\underline{\lim}_{n\to\infty} (a_n + b_n) < \underline{\lim}_{n\to\infty} a_n \vee \overline{\lim}_{n\to\infty} b_n .$$

Thus (56) involves (51).

If $\underline{\lim}_{n\to\infty} a_n$, $\underline{\lim}_{n\to\infty} b_n$, and $\overline{\lim}_{n\to\infty} (a_n + b_n)$ exist then $\underline{\lim}_{n\to\infty} a_n + \underline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n$ and $\underline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n + \underline{\lim}_{n\to\infty} b_n$ are, by Theorem 104, interelements of the sequences $a_n + a_nb_n$ and $a_nb_n + b_n$ respectively. Hence we have, according to the second relation of Theorem 96,

$$(\underline{\lim}_{n\to\infty} a_n + \underline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n) \vee (\underline{\lim}_{n\to\infty} a_n \cdot \underline{\lim}_{n\to\infty} b_n + \underline{\lim}_{n\to\infty} b_n)$$
$$< \overline{\lim}_{n\to\infty} [(a_n + a_nb_n) \vee (a_nb_n + b_n)],$$

i.e. Relation (52). Relation (53) we can prove in the same way, using Theorem 102 instead of Theorem 104.

Theorem 109. *If $\overline{\lim}_{n\to\infty} a_n$ and $\lim_{n\to\infty} b_n$ exist then $\overline{\lim}_{n\to\infty} a_nb_n$ exists too and is equal to $\overline{\lim}_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$.*

Proof. See Theorems 98 and 105.

Theorem 110. *If $\underline{\lim}_{n\to\infty} a_n$ and $\lim_{n\to\infty} b_n$ exist then $\underline{\lim}_{n\to\infty} (a_n \vee b_n)$ exists too and is equal to $\underline{\lim}_{n\to\infty} a_n \vee \lim_{n\to\infty} b_n$.*

Proof. See Theorems 100 and 106.

Theorem 111. *If* $\lim_{n\to\infty} a_n$ *and* $\varliminf_{n\to\infty} b_n$ *exist then* $\varlimsup_{n\to\infty} (a_n + a_nb_n)$ *exists too and* is equal to $\lim_{n\to\infty} a_n + \lim_{n\to\infty} a_n \cdot \varliminf_{n\to\infty} b_n$.

Proof. See Theorems 104 and 107.

Theorem 112. *If* $\varlimsup_{n\to\infty} a_n$ *and* $\lim_{n\to\infty} b_n$ *exist then* $\varlimsup_{n\to\infty} (a_n + a_nb_n)$ *exists too and is equal to* $\varlimsup_{n\to\infty} a_n + \varlimsup_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$.

Proof. See Theorems 102 and 107.

Theorem 113. *If* $\varliminf_{n\to\infty} a_n$ *and* $\lim_{n\to\infty} b_n$ *exists then* $\varliminf_{n\to\infty} a_n + \lim_{n\to\infty} b_n$ *is an interelement of the sequence* $a_n + b_n$.

Proof. We have, by Theorem 95,

$$\varliminf_{n\to\infty} (a_n + a_n b_n) = \varliminf_{n\to\infty} a_n + \varliminf_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n \tag{57}$$

and, by Theorem 111,

$$\varlimsup_{n\to\infty} (a_n b_n + b_n) = \varliminf_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n + \lim_{n\to\infty} b_n . \tag{58}$$

From (57) and (58) and from Theorem 100 follows Theorem 113.

Theorem 114. *If* $\lim_{n\to\infty} a_n$ *and* $\varlimsup_{n\to\infty} b_n$ *exist then* $\lim_{n\to\infty} a_n + \varlimsup_{n\to\infty} b_n$ *is an interelement of the sequence* $a_n + b_n$.

Proof. Compare the proof of Theorem 113 and see Theorems 95, 112, and 100.

Theorem 115. *If the sequences* a_n *and* b_n *converge, the sequences* a_nb_n , $a_n \vee b_n$, *and* $a_n + b_n$ *converge too, and we have*

$$\lim_{n\to\infty} a_n b_n = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n , \tag{59}$$

$$\lim_{n\to\infty} (a_n \vee b_n) = \lim_{n\to\infty} a_n \vee \lim_{n\to\infty} b_n , \text{ and} \tag{60}$$

$$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n . \tag{61}$$

Proof. (59) follows from Theorems 93 and 109, and (60) from Theorems 94 and 110. Further Theorems 95 and 111 (or 112) involve

$$\lim_{n\to\infty} (a_n + a_n b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n \tag{62}$$

and

$$\lim_{n\to\infty} (a_n b_n + b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n + \lim_{n\to\infty} b_n . \tag{63}$$

From (62) and (63) follows (61), by the aid of (60).

Theorem 116. *If* $a_k a_l = 0$ *for* $k, l = 1, 2, 3, \cdots , k \neq l$ *then*

$$\lim_{n\to\infty} a_n = 0.$$

Proof. 0 is obviously the only subelement of the sequence a_n ; hence we have, by Theorem 21,

$$\underline{\lim}_{n \to \infty} a_n = 0.$$

If a is an arbitrary interelement of the sequence a_n , then aa_k is an interelement of the sequence $a_n a_k$ for every k. Hence $aa_k = 0$ for $k = 1, 2, 3, \cdots$. (See Theorem 25.) Since a is also an interelement of the sequence aa_n , we obtain

$$a = 0.$$

This proves the theorem.

7. The Closure-Topology $\Gamma(\tau(A))$

The sequential topology $\tau(A)$ in a Boolean ring A, defined by Definition 8, determines a closure-topology $\Gamma(\tau(A))$. (Compare (III), Theorem 15.) Our Theorem 29 involves that a subclass of A is closed if and only if it contains with every convergent sequence also its limit.

Theorem 117. *Any closed subring $\mathfrak{a}$ of a Boolean σ-ring A is itself a Boolean σ-ring.*

Proof. If a_n $(n = 1, 2, 3, \cdots)$ are elements of $\mathfrak{a}$ then $\prod_{n=1}^{\infty}{}^{(A)} a_n$ and $\sum_{n=1}^{\infty}{}^{(A)} a_n$ exist and are contained in $\mathfrak{a}$. Hence, by Theorem 47, $\prod_{n=1}^{\infty}{}^{(\mathfrak{a})} a_n$ and $\sum_{n=1}^{\infty}{}^{(\mathfrak{a})} a_n$ exist too. (They are even identical with $\prod_{n=1}^{\infty}{}^{(A)} a_n$ and $\sum_{n=1}^{\infty}{}^{(A)} a_n$ respectively.)

Theorem 118. *If $\mathfrak{a}$ is an invariant σ-subring of a Boolean ring A, and a_n $(n = 1, 2, 3, \cdots)$ are elements of $\mathfrak{a}$, then ${}^{(A)}\underline{\lim}_{n \to \infty} a_n$ and ${}^{(A)}\overline{\lim}_{n \to \infty} a_n$ exist and are contained in $\mathfrak{a}$. Hence $\mathfrak{a}$ is especially a closed subring of A.*

Proof. See Theorems 38, 51, and 52.

Theorem 119. *If a_n $(n = 1, 2, 3, \cdots)$ are elements of a closed ideal $\mathfrak{a}$ of a Boolean ring A, and a is an interelement of the sequence a_n with respect to A, then a is contained in $\mathfrak{a}$.*

Proof. We have, by Theorem 42,

$${}^{(A)}\lim_{n \to \infty} \sum_{k=1}^{n}{}^{(A)} a_k a = \sum_{k=1}^{\infty}{}^{(A)} a_k a = a.$$

$a_k a$ $(k = 1, 2, 3, \cdots)$ are elements of $\mathfrak{a}$ because $\mathfrak{a}$ is an ideal. The same holds also for all elements of the sequence $\sum_{k=1}^{n}{}^{(A)} a_k a$ $(n = 1, 2, 3, \cdots)$; thus also for its limit a because $\mathfrak{a}$ is closed.

Theorem 120. *Every normal ideal of a Boolean ring is closed.*

Theorem 120 can be easily deduced from the definition of a normal ideal and from Lemma 5.

THEOREM 121. *If A is an invariant subring of a Boolean ring B, N is the maximal join-extension of A invariant in B, a_n $(n = 1, 2, 3, \cdots)$ are elements of N, a is an element of B, and either of the relations ${}^{(B)}\underline{\lim}_{n\to\infty}\, a_n = a$ and $\overline{\lim}^{(B)}_{n\to\infty}\, a_n = a$ is valid, then a is contained in N. N is especially a closed subring of B.*

PROOF. Let $\bar{B}$ be a complete invariant extension of B (see Theorem 69) so that A is, by Theorem 53, invariant in $\bar{B}$, and $\bar{N}$ the maximal join-extension of A invariant in $\bar{B}$. (Compare the proof of Theorem 70.) If now a_n $(n = 1, 2, 3, \cdots)$ are elements of N, and ${}^{(B)}\underline{\lim}_{n\to\infty}\, a_n$ exists, we have, by Theorem 51,

$$ {}^{(\bar{B})}\underline{\lim}_{n\to\infty}\, a_n = {}^{(B)}\underline{\lim}_{n\to\infty}\, a_n ; $$

on the other side the elements a_n are also elements of $\bar{N}$. Hence ${}^{(\bar{B})}\underline{\lim}_{n\to\infty}\, a_n$ is, by Theorems 70, 75, and 118, an element of $\bar{N}$. From this and from the definitions of N and $\bar{N}$ it can be easily concluded that ${}^{(B)}\underline{\lim}_{n\to\infty}\, a_n$ is an element of N. If a_n $(n = 1, 2, 3, \cdots)$ are elements of N, and ${}^{(B)}\overline{\lim}_{n\to\infty}\, a_n$ exists, we can conclude in a similar way.

THEOREM 122. *If A is an invariant subring of a Boolean σ-ring B, the maximal join-extension of A invariant in B is itself a Boolean σ-ring.*

PROOF. See Theorems 117 and 121.

8. Double Sequences

DEFINITION 13. *Suppose that to every element of some class Θ of ordered couples (m, n) of natural numbers there is assigned an element a_{mn} of a given Boolean ring A. We say that such a correspondence defines a double sequence in A if to every natural number k there exist two natural numbers m and n such that $m \geqq k$, $n \geqq k$, and $(m, n) \in \Theta$.*

DEFINITION 14. *Let $a^{(1)}_{mn}$ and $a^{(2)}_{mn}$ be two double sequences in a Boolean ring, and $\Theta^{(1)}$ and $\Theta^{(2)}$ the classes of ordered couples of natural numbers for which they are defined. In this case we say the double sequence $a^{(1)}_{mn}$ to be a partial sequence of the the double sequence $a^{(2)}_{mn}$ if $\Theta^{(1)} \subset \Theta^{(2)}$ and $a^{(1)}_{mn} = a^{(2)}_{mn}$ for $(m, n) \in \Theta^{(1)}$.*

DEFINITION 15. *Let a_{mn} be a double sequence in a Boolean ring A, Θ the class of ordered couples of natural numbers belonging to it, and u an element of A. In this case we call u a subelement of the double sequence a_{mn} if there is a natural number k such that*

$$ u < a_{mn} $$

for $m \geqq k$, $n \geqq k$, $(m, n) \in \Theta$.

DEFINITION 16. *If the signs A, a_{mn}, and Θ have the same meaning as in Definition 15, we call an element o of A a superelement of the double sequence a_{mn} if there is a natural number k such that*

$$ o > a_{mn} $$

for $m \geqq k$, $n \geqq k$, $(m, n) \in \Theta$.

Now the notions of interelement, lower limit, upper limit, convergence, and limit can be defined for double sequences just so as for simple sequences. If a_{mn} is a double sequence in a Boolean ring A, the sign Θ has the same meaning as in Definition 13, and the lower limit of the double sequence a_{mn} with respect to A exists, we denote this lower limit by ${}^{(A)}\underline{\lim}_{\substack{m,n\to\infty\\(m,n)\in\Theta}} a_{mn}$ or more simply by ${}^{(A)}\underline{\lim}_{m,n\to\infty} a_{mn}$ in case Θ is the class of all ordered couples of natural numbers. The signs ${}^{(A)}\overline{\lim}_{\substack{m,n\to\infty\\(m,n)\in\Theta}} a_{mn}$, ${}^{(A)}\overline{\lim}_{m,n\to\infty} a_{mn}$, ${}^{(A)}\lim_{\substack{m,n\to\infty\\(m,n)\in\Theta}} a_{mn}$ and ${}^{(A)}\lim_{m,n\to\infty} a_{mn}$ are to be understood in similar sense. In all these signs the superscript (A) may be omitted if there is no doubt to which Boolean ring they refer.

All theorems on simple sequences obtained until now, except Theorems 31, 32, 33, 34, and 35, can be pronounced with little changes of the wording also for doubles sequences. For instance Theorems 29 and 30 assume the following forms:

THEOREM 29′: *If $a_{mn}^{(1)}$ is a partial sequence of a double sequence $a_{mn}^{(2)}$, $\Theta^{(1)}$ and $\Theta^{(2)}$ are the corresponding classes of ordered couples of natural numbers, and* $\lim_{\substack{m,n\to\infty\\(m,n)\in\Theta^{(2)}}} a_{mn}^{(2)}$ *exists, then* $\lim_{\substack{m,n\to\infty\\(m,n)\in\Theta^{(1)}}} a_{mn}^{(1)}$ *exists too and is equal to* $\lim_{\substack{m,n\to\infty\\(m,n)\in\Theta^{(2)}}} a_{mn}^{(2)}$.

THEOREM 30′. *Any system of a finite number of convergent double sequences with the same limit is again convergent with the same limit, if by a subelement (superelement) of a system of double sequences be meant a common subelement (superelement) of these double sequences, and the notions of an interelement, lower limit, and limit of a system of double sequences be defined accordingly.*

In Theorems 36 and 37 the signs $\prod_{k=n}^{\infty} a_k$ and $\sum_{k=n}^{\infty} a_k$ are to be replaced by the signs $\prod_{\substack{k,l=n\\(k,l)\in\Theta}}^{\infty} a_{kl}$ and $\sum_{\substack{k,l=n\\(k,l)\in\Theta}}^{\infty} a_{kl}$ respectively. Also Theorems 39, 40, 41, and 42 are to be changed in a similar way.

A special case of Theorem 29′ is

THEOREM 123. *If a_{mn} $(m, n = 1, 2, 3, \cdots)$ are elements of a Boolean ring, m_k and n_k $(k = 1, 2, 3, \cdots)$ natural numbers such that $m_k \to \infty$ and $n_k \to \infty$ for $k \to \infty$, and* $\lim_{m,n\to\infty} a_{mn}$ *exists, then* $\lim_{k\to\infty} a_{m_k n_k}$ *exists too and is equal to* $\lim_{m,n\to\infty} a_{mn}$.

It is an open question whether Theorem 123 can be inverted, i.e. whether $\lim_{k\to\infty} a_{m_k n_k} = a$ whenever $m_k \to \infty$ and $n_k \to \infty$ for $k \to \infty$, implies $\lim_{m,n\to\infty} a_{mn} = a$. We shall return to this point once more in §10.

9. Fundamental Sequences

DEFINITION 17. *A sequence of elements a_n of a Boolean ring A is said to be a fundamental sequence with respect to A if*

$${}^{(A)}\lim (a_m + a_n) = 0.$$

THEOREM 124. *A sequence a_n is a fundamental sequence if and only if the sequence a_nb is a fundamental sequence for every element b.*

PROOF. See Definition 17, Theorem 5, and Lemma 5.

THEOREM 125. *A bounded sequence is a fundamental sequence if and only if the following condition is satisfied: if x is an element of the property that*

$$x < o + u$$

whenever u is a subelement, and o a superelement of the sequence, then $x = 0$.

PROOF. If u is a subelement and o a superelement of the sequence a_n then $o + u$ is a superelement of the double sequence $a_m + a_n$. Conversely to every superelement o^* of the double sequence $a_m + a_n$ there exist a subelement u and a superelement o of the sequence a_n such that

$$o^* = o + u. \tag{64}$$

For if o^* is a superelement of the double sequence $a_m + a_n$, there exists a natural number k such that

$$a_m + a_n < o^* \tag{65}$$

for $m \geqq k$, $n \geqq k$. From (65) follows

$$a_m + a_m o^* < a_n < a_m \vee o^*.$$

Hence $a_m + a_m o^*$ is a subelement, and $a_m \vee o^*$ a superelement of the sequence a_n, provided that $m \geqq k$. On the other side (64) is satisfied if we put $u = a_m + a_m o^*$, $o = a_m \vee o^*$.

After this is settled Theorem 125 follows easily from Theorem 22.

It is obvious that the fundamental sequences of a Boolean ring can be defined by their properties stated in Theorems 124 and 125. Compare Definitions 3 and 4.

THEOREM 126. *Any fundamental sequence has at the utmost one interelement.*

PROOF. If a and a^* are interelements of the sequence a_n then $a + aa^*$ and 0 are interelements of the bounded sequence $a_n(a + aa^*)$. Hence 0 is the only subelement of the latter sequence, and if o is a superelement of this sequence, then

$$a + aa^* < o.$$

The more we may say that

$$a + aa^* < o + u$$

whenever u is a subelement and o a superelement of the sequence $a_n(a + aa^*)$. If the sequence a_n is a fundamental sequence, the sequence $a_n(a + aa^*)$ is, by Theorem 124, likewise a fundamental sequence, and from Theorem 125 follows $a + aa^* = 0$. Just so we can prove that $a^* + aa^* = 0$. Hence $a^* = a$.

THEOREM 127. *Every convergent sequence is a fundamental sequence.*

PROOF. See Definition 17 and Relation (61) of Theorem 115.

On the other hand it is not true that every sequence without an interelement is a fundamental sequence. Let A be the Boolean ring of all finite classes of natural numbers, for $n = 1, 2, 3, \cdots$ $a_{2n-1} = \{1, 2, 3, \cdots, n\}$ (the class of the numbers $1, 2, 3, \cdots, n$), $a_2 = 0$ (the void class), and $(n = 2, 3, \cdots)$ $a_{2n} = \{2, 3, \cdots, n\}$. Then the sequence a_n $(n = 1, 2, 3, \cdots)$ has no interelement (there is no element a satisfying $a > u$ for all subelements u of the sequence), and it is not a fundamental sequence either.

THEOREM 128. *If $\mathfrak{a}$ is an invariant subring of a Boolean ring A, and a_n $(n = 1, 2, 3, \cdots)$ are elements of $\mathfrak{a}$, then the sequence a_n is a fundamental sequence with respect to $\mathfrak{a}$ if and only if it is a fundamental sequence with respect to A.*

PROOF. See Theorem 50 and Definition 17.

THEOREM 129. *A sequence of elements a_n of a Boolean ring A is a fundamental sequence with respect to A if and only if the sequence $\mathfrak{a}(a_n)$ is convergent with respect to $\mathfrak{N}$.*

PROOF. See Theorems 38, 68, 126, 127, and 128.

THEOREM 130. *Every monotonic sequence of elements of a Boolean ring is a fundamental sequence.*

PROOF. See Theorems 35 and 129.

DEFINITION 18. *A Boolean ring A is called σ-complete if every fundamental sequence in A converges with respect to A.*

THEOREM 131. *A Boolean ring A is σ-complete if and only if A is a Boolean σ-ring.*

PROOF. If all fundamental sequences of A converge, and a_n $(n = 1, 2, 3, \cdots)$ are elements of A, then from Theorems 32, 34, and 130 follows that $\prod_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} a_n$ exist. If A is a Boolean σ-ring, every fundamental sequence of A has, by Theorems 38 and 126, exactly one interelement.

D. van Dantzig (I) has shown how every topological ring, satisfying the neighbourhood-axioms of Fréchet-Hausdorff and having the property that sum, difference, and product of two elements of the ring are continuous functions of these elements, can be extended to a topological ring of the same kind and of the property that every fundamental sequence of the extended ring is convergent. In this paper we have defined for every Boolean ring A an intrinsic sequential topology $\tau(A)$, and have not considered any other sequential topology in A until now. Therefore if we consider an extension B of A we are interested only in the case that B is again a Boolean ring, and that the sequential topology $\tau(B)$ is an extension of the sequential topology $\tau(A)$. We know already that $\tau(B)$ is certainly an extension of $\tau(A)$ if B is an invariant extension of A. (Theorem 50.) And Theorems 69 and 131 show that any Boolean ring A can be invariantly extended to a Boolean ring B in which every fundamental sequence converges. Now we shall try to define the notion of a *minimal* invariant extension of A of this property, analogous to the notion of a minimal complete invariant extension of A, defined by Definition 12.

THEOREM 132. *If we mean by the sum of two sequences a_n and b_n in a Boolean ring A the sequence $a_n + b_n$, and by their product the sequence a_nb_n, the sequences*

in A form a Boolean ring S; the fundamental sequences in A form a subring F of S, the convergent sequences in A form a subring C of F, and the null sequences (sequences converging towards zero) *form an ideal $\mathfrak{n}$ in C.*

Proof. The first proposition is evident. The second proposition follows from the equations

$$(a_m + b_m) + (a_n + b_n) = (a_m + a_n) + (b_m + b_n)$$

$$a_m b_m + a_n b_n = a_m(b_m + b_n) + (a_m + a_n)b_n$$

and from Theorems 92 and 115, the third from Theorem 115, and the fourth from Theorems 92 and 115.

Now let A_1 be a homomorphic image of F such that there exists a homomorphism γ from F to A_1 determining exactly the ideal $\mathfrak{n}$ of F. (I.e. $\mathfrak{n}$ shall be the class of the elements of F carried by γ into the zero of A_1.) Such a homomorphic image of F is for instance the quotient-ring $F/\mathfrak{n}$. Two elements of C are carried by γ into the same element of A_1 if and only if they are sequences convergent with respect to A and having the same limit with respect to A. Hence if we denote the image of C in A_1 by C^* and assign to every element a of A that element of C^* in which the convergent sequence, having all terms equal to a, is carried by γ, then this correspondence is an isomorphism between A and C^*. Hence A_1 can be chosen in such a way that the image of the sequence $\{a, a, a, \cdots\}$ in A_1 is identical with a itself. In this case A_1 is an extension of A.

Definition 19. *Let the signs A, F, C, and $\mathfrak{n}$ have the same meaning as in Theorem* 132, *and let A_1 be a homomorphic image of F, and γ a homomorphism $F \to A_1$ determining exactly the ideal $\mathfrak{n}$ of F and having the property that if a is an element of A, the element of A_1 in which γ carries the sequence having all terms equal to a, is identical with a. In this case we call A_1 a first fundamental extension of A.*

Two fundamental sequences a_n and b_n have the same image in A_1 if and only if their difference (this is here the same as $a_n + b_n$) is a null sequence. D. van Dantzig calls such fundamental sequences *concurrent.*

Lemma 10. *If A_1 is a first fundamental extension of a Boolean ring A, then A is invariant in A_1.*

Proof. Let $\mathfrak{B}$ be a subclass of A, and b an element of A such that

$$\sum_{a \in \mathfrak{B}}^{(A)} a = b.$$

Further let f be an element of A_1 such that

$$f > a$$

for all elements a of $\mathfrak{B}$. If a_n is a fundamental sequence in A, of which the image in A_1 is f, then the sequence $a_n a + a$ must be contained in $\mathfrak{n}$, i.e. we must have

$$^{(A)}\lim_{n \to \infty} a_n a = a.$$

Hence we have, by Theorem 41,

$$\sum_{k=n}^{\infty}{}^{(A)} a_k a = a$$

and

$$\sum_{a \in \mathfrak{B}}{}^{(A)} \sum_{k=n}^{\infty}{}^{(A)} a_k a = b$$

for $n = 1, 2, 3, \cdots$. Since the joins $\sum_{a \in \mathfrak{B}}^{(A)} a_k a$ exist (see Theorem 13) we can easily conclude that also

$$\sum_{k=n}^{\infty}{}^{(A)} \sum_{a \in \mathfrak{B}}{}^{(A)} a_k a = b.$$

Hence we have

$$\sum_{k=n}^{\infty}{}^{(A)} a_k b = b$$

and, by Theorem 41,

$$^{(A)}\overline{\lim_{n \to \infty}}\, a_n b = b,$$

whence

$$^{(A)}\lim_{n \to \infty} a_n b = b$$

because $a_n b$ is a fundamental sequence. Thus we have $fb = b$ or $f > b$. This proves that

$$\sum_{a \in \mathfrak{B}}{}^{(A_1)} a = b.$$

THEOREM 133. *If the signs A, F, γ, and A_1 have the same meaning as in Definition* 19 *then any fundamental sequence in A converges with respect to A_1 to that element of A_1 which is its image by virtue of γ.*

PROOF. First we consider a monotonic sequence in A. Let $a_n (n = 1, 2, 3, \cdots)$ be elements of A, and for instance $a_k < a_l$ for $k < l$, $k, l = 1, 2, 3, \cdots$. We have

$$^{(A)}\lim_{n \to \infty} a_n a_k = a_k. \qquad (k = 1, 2, 3, \cdots.)$$

Hence we have, if f is the image of the fundamental sequence a_n, for $k = 1, 2, 3, \cdots$

$$fa_k = a_k$$

or

$$f > a_k.$$

If g is an element of A_1, b_n a fundamental sequence in A the image of which is g, and $g > a_k$ for $k = 1, 2, 3, \cdots$, we have for $k = 1, 2, 3, \cdots$

$$^{(A)}\lim_{n\to\infty} (a_k + a_k b_n) = 0,$$

whence

$$\mathfrak{a}(a_k) + \mathfrak{a}(a_k)\, ^{(\mathfrak{R})}\lim_{n\to\infty} \mathfrak{a}(b_n) = 0,$$

$$^{(\mathfrak{R})}\lim_{n\to\infty} \mathfrak{a}(a_n) + {}^{(\mathfrak{R})}\lim_{n\to\infty} \mathfrak{a}(a_n) . {}^{(\mathfrak{R})}\lim_{n\to\infty} \mathfrak{a}(b_n) = 0,$$

$$^{(A)}\lim_{n\to\infty} (a_n + a_n b_n) = 0,$$

and

$$g > f.$$

This proves

$$\sum_{k=1}^{\infty}{}^{(A_1)} a_k = {}^{(A_1)}\lim_{n\to\infty} a_n = f.$$

If $a_k > a_l$ for $k < l$, $k, l = 1, 2, 3, \cdots$, and f is again the image of the fundamental sequence a_n, we can prove in a similar way that

$$\prod_{k=1}^{\infty}{}^{(A_1)} a_k = {}^{(A_1)}\lim_{n\to\infty} a_n = f.$$

Now let a_n be an arbitrary fundamental sequence in A, and f its image. From what we have just proved follows that $\prod_{l=1}^{\infty}{}^{(A_1)} a_{k+l-1}$ and $\sum_{l=1}^{\infty}{}^{(A_1)} a_{k+l-1}$ exist and are the images of the fundamental sequences $\prod_{l=1}^{n} a_{k+l-1}$ ($n = 1, 2, 3, \cdots$) and $\sum_{l=1}^{n} a_{k+l-1}$ ($n = 1, 2, 3, \cdots$) respectively. But we have

$$^{(A)}\lim_{n\to\infty} \left(a_n \prod_{l=1}^{n} a_{k+l-1} + \prod_{l=1}^{n} a_{k+l-1} \right) = 0$$

and

$$^{(A)}\lim_{n\to\infty} \left(a_n + a_n \sum_{l=1}^{n} a_{k+l-1} \right) = 0.$$

Hence

$$\prod_{l=1}^{\infty}{}^{(A_1)} a_{k+l-1} < f < \sum_{l=1}^{\infty}{}^{(A_1)} a_{k+l-1} .$$

From this relation it can be easily concluded that f is an interelement of the sequence a_n with respect to A_1. Since this sequence is, by Theorem 128 and Lemma 10, also with respect to A_1 a fundamental sequence, Theorem 133 is proved. (See Theorem 126.)

LEMMA 11. *If A_1 is a first fundamental extension of a Boolean ring A, then A_1 is also a join-extension of A.*

PROOF. A_1 is, by Lemma 10, an invariant extension of A. Let N be the maximal join-extension of A invariant in A_1. Then we have, by Theorem 121,

$$N \supset A_1$$

because all elements of A_1 are, by Theorem 133, limits of elements of A and by this limits of elements of N with respect to A_1. Hence $N = A_1$.

In the paper (I) the ring which corresponds to the ring denoted here by A_1, has already the property that all its fundamental sequences converge. We shall see that in our case this assertion is in general not true though D. van Dantzig's "Ringkomplettierungsaxiom" (the product of a null sequence and a fundamental sequence is a null sequence) is satisfied, according to Theorem 92.

LEMMA 12. *If A and A^* are Boolean rings between which there exists an isomorphism Λ, and A_1 and A_1^* are first fundamental extensions of A and A^* respectively, then A_1 and A_1^* are isomorphic too, and the isomorphism Λ can be extended to an isomorphism Λ_1 between A_1 and A_1^* in exactly one way.*

PROOF. Λ_1 can be defined in the following way: let f be an element of A_1, a_n a sequence of elements of A converging towards f with respect to A_1, a_n^* the sequence in A^* corresponding to the sequence a_n by virtue of Λ, and

$$f^* = {}^{(A_1^*)}\lim_{n\to\infty} a_n^* ;$$

this correspondence $f \to f^*$ is the desired isomorphism Λ_1.

In the following we shall denote by ρ the smallest ordinal number which is the ordinal type of an uncountable set. More precisely let ψ be the ordinal number of some uncountable well-ordered set, and ρ the smallest ordinal number satisfying $\rho \leqq \psi$ which is as well the ordinal type of an uncountable set.

DEFINITION 20. *Let A be an arbitrary Boolean ring, and α an ordinal number satisfying $\alpha \leqq \rho$. Let us assign a Boolean ring A_μ to any ordinal number μ satisfying $\mu \leqq \alpha$ by the following transfinite induction:*

1.) *A_1 is a first fundamental extension of A;*

2.) *if the ordinal number μ has a predecessor ν (thus $\mu = \nu + 1$) then A_μ is a first fundamental extension of A_ν;*

3.) *if the ordinal number μ is a limit-number then*

$$A_\mu = \sum_{\nu<\mu} A_\nu ,$$

the sign Σ denoting the set-theoretical union; the ring-operations in A_μ are defined in the following way: if a and b are elements of A_μ, and ν an ordinal number such that $\nu < \mu$, $a \in A_\nu$, $b \in A_\nu$, then

$$a\Delta_{A_\mu} b = a\Delta_{A_\nu} b,$$

$$a \wedge_{A_\mu} b = a \wedge_{A_\nu} b.$$

In this case we call A_α a fundamental extension of the Boolean ring A of the order α.

If μ is an ordinal number satisfying $\mu < \alpha$, the Boolean ring A_μ appearing in Definition 20 is obviously a fundamental extension of A of the order μ.

THEOREM 134. *If A is an arbitrary Boolean ring, α an ordinal number satisfying $\alpha \leqq \rho$, and A_α a fundamental extension of A of the order α, then A_α is a join-extension of A.*

PROOF. See Theorems 62 and 63, Lemma 11, and Definition 20.

THEOREM 135. *If A is an arbitrary Boolean ring, α and β are ordinal numbers satisfying $\beta < \alpha \leqq \rho$, A_α and A_β fundamental extensions of A of the orders α and β respectively, and $A_\beta \subset A_\alpha$, then A_α is a join-extension of A_β.*

PROOF. See Theorems 64 and 134.

LEMMA 13. *Let A be a subring of a Boolean ring B, and $A_{(1)}$ the class of all elements of B which are limits of sequences of elements of A with respect to B. Then $A_{(1)}$ is a subring of B containing A. If A is especially invariant in B then $A_{(1)}$ is a join-extension of A invariant in B.*

PROOF. It is evident that $A_{(1)} \supset A$. That $A_{(1)}$ is a subring of B follows from Theorem 115. If A is invariant in B, Theorem 121 involves that $A_{(1)}$ is contained in the maximal join-extension of A invariant in B.

DEFINITION 21. *If the signs A, B, and $A_{(1)}$ have the same meaning as in Lemma 13 then $A_{(1)}$ is called the first limit-ring of A in B.*

DEFINITION 22. *Let A be a subring of a Boolean ring B, and α an ordinal number satisfying $\alpha \leqq \rho$. Let us assign a subring $A_{(\mu)}$ of B to any ordinal number μ satisfying $\mu \leqq \alpha$ by the following transfinite induction:*

1.) *$A_{(1)}$ is the first limit-ring of A in B;*

2.) *if the ordinal number μ has a predecessor ν (so that $\mu = \nu + 1$) then $A_{(\mu)}$ is the first limit-ring of $A_{(\nu)}$ in B;*

3.) *if the ordinal number μ is a limit-number then*

$$A_{(\mu)} = \sum_{\nu<\mu}^{(\mathfrak{R})} A_{(\nu)},$$

by $\mathfrak{R}$ denoted the lattice of all subrings of B.

In this case we call $A_{(\alpha)}$ the limit-ring of A in B of the order α.

THEOREM 136. *If A is an invariant subring of a Boolean ring B, then any limit-ring $A_{(\alpha)}$ of A in B is invariant in B and a join-extension of A and of every limit-ring of A in B of lower order.*

PROOF. From Definition 22 it can be easily concluded that $A_{(\alpha)}$ is contained in the maximal join-extension of A invariant in B. For the rest see Theorems 53, 61, 64, and 70.

THEOREM 137. *If A is any Boolean ring, α an ordinal number satisfying $\alpha \leqq \rho$, A_α a fundamental extension of A of the order α, and B an invariant extension of A_α, then A_α is the limit-ring of A in B of the order α.*

PROOF. Assign a Boolean ring A_μ to any ordinal number μ satisfying $\mu < \alpha$ in such a way that the rings A_μ $(\mu \leqq \alpha)$ satisfy the conditions stated in Definition 20. Let us suppose that $\mu_0 < \alpha$, and that we have already proved that A_{μ_0} is

the limit-ring of A in B of the order μ_0. Then let a_n be a sequence of elements of A_{μ_0} which has a limit a with respect to B. Since A_{μ_0} is, by Theorems 53, 61, and 135, invariant in B, the sequence a_n is a fundamental sequence with respect to A_{μ_0} according to Theorems 127 and 128. Hence this sequence has, by Definitions 19 and 20 and by Theorem 133, a limit with respect to A_{μ_0+1}, and this limit is necessarily identical with a, because also A_{μ_0+1} is invariant in B. Hence $a \epsilon A_{\mu_0+1}$. If we assume conversely that a is an element of A_{μ_0+1} then a is, by Definitions 19 and 20 and Theorem 133, the limit of some sequence of elements of A_{μ_0} with respect to A_{μ_0+1}, and consequently also with respect to B because A_{μ_0+1} is invariant in B. Thus we find that a is an element of the limit-ring of A in B of the order $\mu_0 + 1$. Hence this limit-ring is identical with A_{μ_0+1}. We see that Theorem 137 can be proved by transfinite induction; for the conclusion from the predecessors of a limit-number to this limit-number itself is still simpler.

THEOREM 138. *If A is any Boolean ring, α and β ordinal numbers satisfying $\beta \leqq \alpha \leqq \rho$, and A_α a fundamental extension of A of the order α, then there exists exactly one subring A_β of A_α which is a fundamental extension of A of the order β.*

PROOF. The existence of A_β follows from Definition 20, the uniqueness from Theorem 137. (A_β is the limit-ring of A in A_α of the order β.)

THEOREM 139. *If A is an invariant subring of a Boolean σ-ring B, α an ordinal number satisfying $\alpha \leqq \rho$, and $A_{(\alpha)}$ the limit-ring of A in B of the order α, then $A_{(\alpha)}$ is a fundamental extension of A of the order α.*

PROOF. If a_n $(n = 1, 2, 3, \cdots)$ are elements of A, and the sequence consisting of these elements is a fundamental sequence with respect to A, it is, by Theorem 128, also a fundamental sequence with respect to B; hence it is, by Theorem 131, convergent with respect to B; its limit is an element of $A_{(1)}$. Conversely any element of $A_{(1)}$ can be obtained in this way. If we assign to every fundamental sequence in A that element of $A_{(1)}$ which is its limit with respect to B we get a homomorphism γ from the Boolean ring F of the fundamental sequences in A to $A_{(1)}$ of the sort demanded in Definition 19. This proves the theorem for the case $\alpha = 1$. The general theorem follows now by transfinite induction.

THEOREM 140. *Let A and A^* be Boolean rings between which there exists an isomorphism Λ, α and β ordinal numbers satisfying $\beta \leqq \alpha \leqq \rho$, A_β a fundamental extension of A of the order β, and A^*_α a fundamental extension of A^* of the order α. Then there exists exactly one subring R^* of A^*_α such that the isomorphism Λ can be extended to an isomorphism Λ_β between A_β and R^*. Λ_β is uniquely determined, and R^* is a fundamental extension of A^* of order β. If especially $\beta = \alpha$ then $R^* = A^*_\alpha$.*

PROOF. Let A^*_β be a subring of A^*_α which is a fundamental extension of A^* of the order β. That A^*_β exists and is unique follows from Theorem 138. If we put $R^* = A^*_\beta$ it ensues from Lemma 12 and by transfinite induction that the isomorphism Λ_β, demanded by Theorem 140, exists and is unique. The uniqueness of R^* is involved by Theorem 83.

THEOREM 141. *If A is a subring of a Boolean ring B the closure of A in B is as well a subring of B; it is identical with the limit-ring $A_{(\rho)}$ of A in B of the order ρ.*

PROOF. Let a_n be a sequence of elements of $A_{(\rho)}$ which has a limit a with re-

spect to B. Further, for $n = 1, 2, 3, \cdots$ let α_n be the smallest ordinal number which is not larger than ρ and has the property that a_n is contained in $A_{(\alpha_n)}$, by $A_{(\beta)}$ ($\beta \leqq \rho$) denoted the limit-ring of A in B of the order β. Since ρ is a limit-number we have certainly

$$\alpha_n \neq \rho$$

for $n = 1, 2, 3, \cdots$. Further let α be the smallest ordinal number satisfying the inequality

$$\alpha_n \leqq \alpha \leqq \rho$$

for $n = 1, 2, 3, \cdots$. We have

$$\alpha \neq \rho;$$

for the asumption $\alpha = \rho$ would imply that ρ would be the ordinal number of a finite or countable set, contrary to the definition of ρ. Hence we have also

$$\alpha + 1 \leqq \rho.$$

Now the elements a_n ($n = 1, 2, 3, \cdots$) are all contained in $A_{(\alpha)}$. Consequently the element

$$a = {}^{(B)}\lim_{n \to \infty} a_n$$

is contained in $A_{(\alpha+1)}$. Thus we have

$$a \in A_{(\rho)}$$

and it is proved that $A_{(\rho)}$ is a closed subclass of B. On the other side it can be shown by transfinite induction that any subclass of B closed under $\Gamma(\tau(B))$ and containing A contains $A_{(\alpha)}$ whenever $\alpha \leqq \rho$, and by this especially $A_{(\rho)}$. Thus Theorem 141 is proved.

If B is especially the Boolean ring of all sets of real numbers, and A the subclass of B defined by the property that a is an element of A if and only if it is either the empty set or the union of a finite number of sets, every one of which is either an open real interval or consists of exactly one real number (see the example stated after Definition 9), then $A_{(\rho)}$ is the Boolean ring of the Borel sets of real numbers. In this case it is known that the rings $A_{(\alpha)}$ ($\alpha \leqq \rho$) are all different. Since B is a complete Boolean ring here, $A_{(\alpha)}$ is, by Theorem 139, a fundamental extension of A of the order α. It follows that if A_1 is a first fundamental extension of A, A_1 is not a σ-complete Boolean ring. If we compare this result with the results of §7 of the paper (I), we find that the sequential topology $\tau(A)$ of a Boolean ring A defined by Definition 8, in general cannot be derivated from a neighbourhood-topology satisfying the axioms of Fréchet-Hausdorff.

THEOREM 142. *If A_ρ is a fundamental extension of a Boolean ring A of the order ρ then A_ρ is a Boolean σ-ring.*

PROOF. Let B be a Boolean σ-ring and an invariant extension of A_ρ. That such a Boolean ring B exists follows from Theorem 69. Then A_ρ is, by Theorem 137, the limit-ring of A in B of the order ρ, and it follows from Theorem 141 that A_ρ is closed in B. Hence A_ρ is a Boolean σ-ring, according to Theorem 117.

DEFINITION 23. *An extension $\tilde{A}$ of a Boolean ring A is said to be a minimal invariant σ-extension of A, or shorter: a σ-completion of A, if $\tilde{A}$ is a Boolean σ-ring invariant over A, and no proper subring of $\tilde{A}$ containing A is itself a Boolean σ-ring.*

THEOREM 143. *An extension $\tilde{A}$ of a Boolean ring A is a minimal invariant σ-extension of A if and only if $\tilde{A}$ is a fundamental extension of A of the order ρ.*

PROOF. If $\tilde{A}$ is a minimal invariant σ-extension of A, the limit-ring of A in $\tilde{A}$ of the order ρ is, by Theorem 139, a fundamental extension of A of the order ρ; hence it is, by Theorem 142, a Boolean σ-ring, and Definition 23 involves that it cannot be different from $\tilde{A}$. If $\tilde{A}$ is a fundamental extension of A of the order ρ then $\tilde{A}$ is, by Theorem 142, a Boolean σ-ring and, by Theorem 134, a join-extension and by this an invariant extension of A. (Theorem 61.) On the other side $\tilde{A}$ is, by Theorems 137 and 141, the closure of A in $\tilde{A}$. If R is a σ-subring of $\tilde{A}$ containing A then R is, by Theorems 61 and 64, invariant in $\tilde{A}$ and, by Theorem 118, closed in $\tilde{A}$. Hence $R = \tilde{A}$, and it is proved that $\tilde{A}$ is a σ-completion of A.

THEOREM 144. *If $\tilde{A}$ is a σ-completion of a Boolean ring A, and B an invariant extension of $\tilde{A}$, then $\tilde{A}$ is the closure of A in B.*

PROOF. See Theorems 137, 141, and 143.

THEOREM 145. *If A is an invariant subring of a Boolean σ-ring B, the closure of A in B is a minimal invariant σ-extension of A.*

PROOF. See Theorems 139, 141, and 143.

THEOREM 146. *If $\tilde{A}$ and $\tilde{A}^*$ are minimal invariant σ-extensions of the same Boolean ring A then $\tilde{A}$ and $\tilde{A}^*$ are isomorphic, and there is exactly one isomorphism between $\tilde{A}$ and $\tilde{A}^*$ carrying every element of A into itself.*

Hence we may say that $\tilde{A}$ and $\tilde{A}^*$ are abstractly identical.

PROOF. See Theorems 140 and 143.

10. The Sequential Topology $\tau(\Gamma(\tau(A)))$

If A is any Boolean ring, the closure-topology $\Gamma(\tau(A))$ in A, considered in §7, determines a new sequential topology $\tau(\Gamma(\tau(A)))$ in A according to (III), Theorem 15. The closure-topology determined by this sequential topology is identical with $\Gamma(\tau(A))$ or, in other words, the closure-topology $\Gamma(\tau(A))$ and the sequential topology $\tau(\Gamma(\tau(A)))$ are equivalent. (See (III), Definition 8 and Theorem 18.)

LEMMA 14. *If a sequence a_n has a limit a under $\tau(A)$ it has also under $\tau(\Gamma(\tau(A)))$ the limit a and no other limit.*

PROOF. It is obvious that the sequence a_n has the limit a also under $\tau(\Gamma(\tau(A)))$. If $b \neq a$, at the utmost a finite number of terms of the sequence a_n is equal to b, and Theorem 29 involves that the terms of this sequence different from b and the element a together constitute a set closed under $\Gamma(\tau(A))$. The

complement of this set is an open set which contains b, but fails to contain infinitely many terms of the sequence a_n. Hence b is certainly no limit of the sequence a_n under $\tau(\Gamma(\tau(A)))$. This completes the proof of Lemma 14.

Further it is obvious that our Theorems 25, 29, and 30 hold also for the sequential topology $\tau(\Gamma(\tau(A)))$. (They hold in any derivative sequential topology.)

THEOREM 147. (Compare (VII), Satz 29.) *If a_n $(n = 1, 2, 3, \cdots)$ and a are elements of a Boolean ring A then a is a limit of the sequence a_n under $\tau(\Gamma(\tau(A)))$ if and only if every partial sequence of the sequence a_n contains a further partial sequence converging towards a under $\tau(A)$. Hence any sequence of elements of A has at the utmost one limit under $\tau(\Gamma(\tau(A)))$.*

PROOF. If a set open under $\Gamma(\tau(A))$ and containing a fails to contain an infinite partial sequence of the given sequence, no partial sequence of the mentioned partial sequence can converge towards a under $\tau(A)$. Hence if every partial sequence of the given sequence contains a further partial sequence converging towards a under $\tau(A)$, the given sequence has certainly the limit a under $\tau(\Gamma(\tau(A)))$.

Let us suppose conversely that the sequence a_n has the limit a under $\tau(\Gamma(\tau(A)))$. If the sequence a_n had not the property asserted in Theorem 147 we could state a partial sequence a_{n_k} $(k = 1, 2, 3, \cdots)$ of this sequence containing no further partial sequence converging towards a under $\tau(A)$. Then the generalized Theorem 29 and Lemma 14 involve that the sequence a_{n_k} contains no sequence convergent under $\tau(A)$ at all. Moreover we may suppose that the elements a_{n_k} are all different from a. Thus the elements of A which are different from all elements a_{n_k} form an open set which contains a, but fails to contain infinitely many terms of the sequence a_n. This result would contradict to the supposition that the sequence a_n converges towards a under $\tau(\Gamma(\tau(A)))$.

DEFINITION 24. *If a_n $(n = 1, 2, 3, \cdots)$ and a are elements of a Boolean ring A, we say that the sequence a_n $(n = 1, 2, 3, \cdots)$ converges weakly towards a with respect to A, or that a is its weak limit with respect to A, and write*

$$^{(A)}\mathrm{Lim}_{n\to\infty}\, a_n = a$$

if the sequence a_n converges towards a under $\tau(\Gamma(\tau(A)))$.

The words "with respect to A" and the superscript in the sign $^{(A)}\mathrm{Lim}_{n\to\infty}\, a_n$ may be omitted if there is no doubt which Boolean ring is meant.

We have chosen the words "weak convergence" and "weak limit" for convenience although the considered convergence under $\tau(\Gamma(\tau(A)))$ is no analogue of the weak convergence in a linear metric space.

As far as the sequential topology $\tau(A)$ is concerned, we shall use the same terminology and denotation in this paragraph as in the preceding ones.

The sequential topology $\tau(\Gamma(\tau(A)))$ is in general different from the sequential topology $\tau(A)$. Let A be the Boolean ring of those sets of real numbers which

are either empty or unions of finite numbers of right-hand open intervals. (A is identical with the Boolean ring $\mathfrak{b}$ considered after Definition 9.) If μ and ν are real numbers we denote the interval $\mu \leqq \xi < \nu$ by $< \mu, \nu)$. Now put

$$a_n = < \frac{n - 2^m}{2^m}, \frac{n - 2^m + 1}{2^m} \Big)$$

for $2^m \leqq n < 2^{m+1}$, $m = 0, 1, 2, 3, \cdots$. a_n $(n = 1, 2, 3, \cdots)$ are elements of A. If m and n are natural numbers, at least one of the equations

$$a_m < a_n,$$

$$a_m > a_n,$$

and

$$a_m a_n = 0$$

is valid. To a given natural number n_0 there exist at the utmost finitely many natural numbers n such that

$$a_n > a_{n_0}.$$

Hence any partial sequence b_n of the sequence a_n contains at least either a partial sequence c_n satisfying

$$(67) \qquad c_{n+1} < c_n \qquad (n = 1, 2, 3, \cdots)$$

or a partial sequence c_n satisfying

$$(68) \qquad c_k c_l = 0 \qquad (k, l = 1, 2, 3, \cdots, k \neq l).$$

(67) as well as (68) implies

$$\lim_{n \to \infty} c_n = 0.$$

(See Theorem 116.) Hence we have

$$\operatorname*{Lim}_{n \to \infty} a_n = 0.$$

On the other side we have

$$\sum_{n=2^m}^{2^{m+1}-1} a_n = a_1 \qquad (m = 1, 2, 3, \cdots),$$

thus

$$\sum_{k=n}^{\infty} a_k = a_1 \qquad (n = 1, 2, 3, \cdots)$$

and

$$\overline{\lim_{n \to \infty}} a_n = a_1.$$

Since $a_1 \neq 0$ the equation $\lim_{n \to \infty} a_n = 0$ is not valid.

THEOREM 148. *If* $\operatorname*{Lim}_{n\to\infty} a_n$ *exists then it is an interelement of the sequence* a_n *and of any partial sequence of this sequence.*

PROOF. See Theorems 27 and 147.

Now it is obvious that our Theorems 50, 87, 92, and 115 are valid also for the weak convergence. The same holds for Theorem 91 if it is pronounced only for convergent sequences. I.e. we have

THEOREM 149. *If* $a_n < b_n$ *for* $n = 1, 2, 3, \cdots$, *and* $\operatorname*{Lim}_{n\to\infty} a_n$ and $\operatorname*{Lim}_{n\to\infty} b_n$ *exist, then*

$$\operatorname*{Lim}_{n\to\infty} a_n < \operatorname*{Lim}_{n\to\infty} b_n .$$

PROOF. We have for $n = 1, 2, 3, \cdots$

$$a_n = a_n b_n ,$$

hence, by the generalized Theorem 115,

$$\operatorname*{Lim}_{n\to\infty} a_n = \operatorname*{Lim}_{n\to\infty} a_n \cdot \operatorname*{Lim}_{n\to\infty} b_n$$

proving Theorem 149.

If a_{mn} is a double sequence in a Boolean ring A, Θ the class of ordered couples of natural numbers for which the sign a_{mn} is defined, and a an element of A, we say that the double sequence a_{mn} converges weakly towards a, or has the weak limit a, and write

$$^{(A)}\operatorname*{Lim}_{\substack{m,n\to\infty\\(m,n)\in\Theta}} a_{mn} = a,$$

if to any open set containing a there can be assigned a natural number k such that a_{mn} is contained in this open set for $m \geqq k$, $n \geqq k$, $(m, n) \in \Theta$. If Θ is the class of all ordered couples of natural numbers, we write simply $^{(A)}\operatorname*{Lim}_{m,n\to\infty} a_{mn}$ instead of $^{(A)}\operatorname*{Lim}_{\substack{m,n\to\infty\\(m,n)\in\Theta}} a_{mn}$. In both cases the superscript (A) may be omitted if there is no doubt which Boolean ring is meant.

Theorems 25, 29′ (8), and 123 hold also for weakly-convergent double sequences. Moreover Theorem 123 pronounced for weakly-convergent double sequences can be inverted. I.e. if $m_k \to \infty$, $n_k \to \infty$ for $k \to \infty$ implies $\operatorname*{Lim}_{k\to\infty} a_{m_k n_k} = a$, then $\operatorname*{Lim}_{m,n\to\infty} a_{mn} = a$. (To prove this proposition is not difficult.) In particular if $m_k \to \infty$, $n_k \to \infty$ for $k \to \infty$ implies $\lim_{k\to\infty} a_{m_k n_k} = a$, we have certainly $\operatorname*{Lim}_{m,n\to\infty} a_{mn} = a$. On the other hand compare the remark after Theorem 123.—Any double sequence of a Boolean ring A has at the utmost one weak limit with respect to A.

Now it is clear that also Theorems 50, 87, 92, 115 and 149 hold for weakly-convergent double sequences.

DEFINITION 25. *A sequence of elements a_n of a Boolean ring A is said to be a weak fundamental sequence with respect to A if*

$$^{(A)}\operatorname*{Lim}_{m,n\to\infty} (a_m + a_n) = 0.$$

Theorems 127, 128, and 132 continue to be right if we replace the notions of "fundamental sequence" and "convergence" wherever they appear, by the notions of "weak fundamental sequence" and "weak convergence."

THEOREM 150. *Any two weakly-concurrent fundamental sequences are concurrent.*

PROOF. If a_n and b_n are fundamental sequences, and

$$\operatorname*{Lim}_{n\to\infty} (a_n + b_n) = 0,$$

then there exists, by Theorem 147, a sequence n_k of natural numbers different from each other such that

$$\lim_{k\to\infty} (a_{n_k} + b_{n_k}) = 0. \tag{69}$$

On the other side we have, by Theorem 123 and Definition 17,

$$\lim_{k\to\infty} (a_k + a_{n_k}) = 0 \tag{70}$$

and

$$\lim_{k\to\infty} (b_k + b_{n_k}) = 0. \tag{71}$$

From (69), (70), and (71) we get, according to Theorem 115,

$$\lim_{k\to\infty} (a_k + b_k) = 0.$$

THEOREM 151. *If b_n is a fundamental sequence, and a_n a sequence of the property that every partial sequence of the sequence a_n contains a fundamental sequence weakly-concurrent with the fundamental sequence b_n, then a_n is a weak fundamental sequence weakly-concurrent with the fundamental sequence b_n.*

PROOF. First we observe that any partial sequence of a fundamental sequence is a fundamental sequence concurrent with the whole sequence. This can be easily concluded from Definition 17 and from Theorems 29′ and 123. Now if we notice Theorems 147 and 150, we find that the suppositions of Theorem 151 involve

$$\operatorname*{Lim}_{n\to\infty} (a_n + b_n) = 0.$$

Since

$$a_m + a_n = (a_m + b_m) + (a_n + b_n) + (b_m + b_n),$$

and

$$\lim_{m,n\to\infty} (b_m + b_n) = 0,$$

we have

$$\lim_{m,n\to\infty} (a_m + a_n) = 0,$$

according to the generalized Theorem 115.

On the other hand it is an open question whether Theorem 151 can be inverted, i.e. whether every weak fundamental sequence of a Boolean ring contains a fundamental sequence in the sense of Definition 17. If this inversion is true, it follows that every weak fundamental sequence of a Boolean σ-ring is weakly-convergent. If it were not true then it could occur that a Boolean ring (if you want even a complete one) could not be invariantly extended to a Boolean ring in which every weak fundamental sequence is weakly-convergent.

PRAGUE

REFERENCES

(I) D. van Dantzig: *Zur topologischen Algebra.* I. Komplettierungstheorie. Mathematische Annalen 107(1933), pp. 587–626.

(II) Garrett Birkhoff: *On the Structure of Abstract Algebras.* Proceedings of the Cambridge Philosophical Society 31(1935), pp. 433–454.

(III) Garrett Birkhoff: *On the Combination of Topologies.* Fundamenta Mathematicae 26(1936), pp. 156–166.

(IV) Fritz Klein: *Boole-Schrödersche Verbände.* Deutsche Mathematik 1 (1936), pp. 528–537.

(V) M. H. Stone: *The Theory of Representations for Boolean Algebras.* Transactions of the American Mathematical Society 40(1936), pp. 37–111.

(VI) Garrett Birkhoff: *Rings of Sets.* Duke Mathematical Journal 3(1937), pp. 443–454.

(VII) L. V. Kantorovitch: *Lineare halbgeordnete Räume.* Recueil Mathématique 2/44(1937), pp. 121–168.

(VIII) Garrett Birkhoff: *Lattices and their Applications.* Bulletin of the American Mathematical Society 44(1938), pp. 793–800.

ANNALS OF MATHEMATICS
Vol. 42, No. 5, December, 1941

ON INCIDENCE MATRICES, NUCLEI AND HOMOTOPY TYPES

BY J. H. C. WHITEHEAD

(Received January 16, 1940)

1. The main object of this paper is to establish as close a connection as possible between certain purely algebraic processes on the one hand, and the theory of nuclei[1] and homotopy types[2] of complexes on the other. The algebraic processes include the transformations, operating on the incidence matrices with elements in the group ring of a given complex, by means of which Reidemeister's invariants are defined.[3] But in the main theorems, namely theorems 8 and 9 in §4, the incidence matrices $\mathbf{r}^n, \cdots, \mathbf{r}^1$ are replaced by what we call a natural system for a given complex $\mathbf{K}^n$. A natural system, $(\mathbf{r}, R)$, consists of the matrices $\mathbf{r}^n, \cdots, \mathbf{r}^3$, together with a "natural" system, R, of generators and relations, for the fundamental group $\pi_1(\mathbf{K}^n)$. Such systems, defined purely algebraically, are classified by two kinds of equivalence, called L-equivalence and L^*-equivalence, of which the former implies the latter. Theorem 8 asserts two kinds of combinatorial invariance. It states that natural systems for two given complexes $\mathbf{K}_0^n$ and $\mathbf{K}_1^n$ are "L-equivalent" if $\mathbf{K}_0^n$ and $\mathbf{K}_1^n$ have the same nucleus, and "L^*-equivalent" if $\mathbf{K}_0^n$ and $\mathbf{K}_1^n$ are of the same homotopy type. Theorem 9 is complementary to theorem 8, and states that the elementary transformations by means of which X-equivalence is defined ($X = L$ or L^*), can be "copied geometrically." More precisely, if $(\mathbf{r}, R)$ is a natural system for $\mathbf{K}^n$, then any system which is X-equivalent to $(\mathbf{r}, R)$, is a natural system for some complex of the same homotopy type as $\mathbf{K}^n$, and for one which has the same nucleus as $\mathbf{K}^n$ if $X = L$. This means that one can carry out certain types of algebraic calculation without destroying the geometrical significance of the system on which one is operating, and the result is a step towards translating the theory of nuclei and m-groups, likewise homotopy types (of polyhedra), into purely algebraic terms.

As an application of the theorems in §§3 and 4 we prove, in §5, that two lens

[1] See J. H. C. Whitehead, Proc. L. M. S., 45 (1939), 243-327, which will be referred to as S.S. Throughout this paper we shall only be interested in the nucleus and homotopy type of a complex as principles of classification.

[2] See pp. 124-5 of W. Hurewicz, Proc. Kon. Akad. Amsterdam, 39 (1936), 117-25, or note III_1 in §6 below. Any two complexes with the same nucleus are of the same homotopy type.

[3] See, among other papers, K. Reidemeister, Abh. Hamb. Sem., 10 (1934), 211-15; 11 (1935), 102-9 and Journal für die r. u. a. Math., 173 (1935), 164-73. By the group ring of a complex we mean the group ring of its fundamental group, with (rational) integral coefficients. We confine ourselves to the incidence matrices which are defined in terms of the universal covering complex, leaving aside the concept of "Ueberdeckungen."

spaces of types[4] (m, q) and (m, r) are of the same homotopy type if they have the same intersection invariant,[5] that is to say if $r \equiv \pm l^2 q \pmod m$, for some value of l. Thus the "Hauptvermutung" would imply the negative answer to the question, proposed by W. Hurewicz,[6] "are two closed manifolds of the same homotopy type necessarily homeomorphic?" In fact two closed simplicial manifolds of the same homotopy type need not even have the same nucleus, which adds interest to the question "is the nucleus of a complex a topological invariant?" This theorem on lens spaces, and a simpler example given in §6, also imply the negative answer to each of the questions Q. 1, $\cdots$, 4, which were left open in S.S., §10.

2. Let $\mathfrak{R}$ be any ring with a unit element e. We recall that an element $a \in \mathfrak{R}$ is said to be *regular* if, and only if, there are elements a' and a'', such that $a'a = aa'' = e$. In this case $a' = a'aa'' = a'' = a^{-1}$, say, and a^{-1} is called the *inverse* of a. The square matrices of a given degree n (i.e. having n rows and columns) with elements in $\mathfrak{R}$, also constitute a ring with a unit element. A regular element in this ring of matrices will be called a *regular matrix*, and we shall denote the inverse of a regular matrix $\mathbf{a}$ by $\mathbf{a}^{-1}$. Let

$$\mathbf{r} = \begin{Vmatrix} \mathbf{a}_1 & \mathbf{b}_1 \\ 0 & \mathbf{b}_2 \end{Vmatrix}$$

be a regular matrix of degree $m + n$, where $\mathbf{a}_1$ is a square matrix of degree m, $\mathbf{b}_2$ is a square matrix of degree n and 0 represents a rectangular array of zeros (i.e. each of its elements is the zero in $\mathfrak{R}$). Further let $\mathbf{b}_2$ have a left inverse, meaning an n-rowed square matrix $\mathbf{c}$, such that $\mathbf{c}\mathbf{b}_2 = \mathbf{e}_n$ where $\mathbf{e}_n$ denotes the unit matrix of degree n.

LEMMA 1. *Under these conditions* $\mathbf{a}_1$ *and* $\mathbf{b}_2$ *are regular matrices and*

$$\mathbf{r}^{-1} = \begin{Vmatrix} \mathbf{a}_1^{-1} & \mathbf{a}_2' \\ 0 & \mathbf{b}_2^{-1} \end{Vmatrix}$$

where $\mathbf{a}_2' = -\mathbf{a}_1^{-1}\mathbf{b}_1\mathbf{b}_2^{-1}$.

Let

$$\mathbf{r}^{-1} = \begin{Vmatrix} \mathbf{a}_1' & \mathbf{a}_2' \\ \mathbf{b}_1' & \mathbf{b}_2' \end{Vmatrix}$$

[4] See H. Seifert and W. Threlfall, Lehrbuch der Topologie, Leipzig (1933), p. 215.

[5] J. W. Alexander, Proc. Nat. Acad. of Sciences, 10 (1924), 99–101; Seifert and Threlfall, loc. cit., p. 279. This contradicts J. H. C. Whitehead, Quarterly Journal of Math. Oxford series, 10 (1939), 81–3, which will be referred to as Q.J. In Q.J. I overlooked the importance of the additional row, which the n^{th} incidence matrix of an n-dimensional complex receives by reason of an elementary expansion of order $n + 1$. This error was pointed out to me by Shaun Wylie.

[6] Loc. cit., p. 125.

where $\mathbf{a}_1'$ and $\mathbf{b}_2'$ are square matrices of degrees m and n respectively. Then the relation $\mathbf{r}\mathbf{r}^{-1} = \mathbf{e}_{m+n}$ implies $\mathbf{a}_1\mathbf{a}_1' + \mathbf{b}_1\mathbf{b}_1' = \mathbf{e}_m$, $\mathbf{a}_1\mathbf{a}_2' + \mathbf{b}_1\mathbf{b}_2' = 0$, $\mathbf{b}_2\mathbf{b}_1' = 0$ and $\mathbf{b}_2\mathbf{b}_2' = \mathbf{e}_n$. Since $\mathbf{c}\mathbf{b}_2 = \mathbf{b}_2\mathbf{b}_2' = \mathbf{e}_n$, it follows that $\mathbf{b}_2$ is regular, that $\mathbf{b}_2' = \mathbf{b}_2^{-1}$, and also that $\mathbf{b}_1' = \mathbf{b}_2^{-1}\mathbf{b}_2\mathbf{b}_1' = 0$. Therefore $\mathbf{a}_1\mathbf{a}_1' = \mathbf{e}_m$, and the relation $\mathbf{r}^{-1}\mathbf{r} = \mathbf{e}_{m+n}$ implies $\mathbf{a}_1'\mathbf{a}_1 = \mathbf{e}_m$, whence $\mathbf{a}_1$ is regular and $\mathbf{a}_1' = \mathbf{a}_1^{-1}$. Finally $\mathbf{a}_2' = -\mathbf{a}_1^{-1}\mathbf{b}_1\mathbf{b}_2' = -\mathbf{a}_1^{-1}\mathbf{b}_1\mathbf{b}_2^{-1}$ and the proof is complete.

Let $\mathfrak{G}$ be a specified sub-group of the multiplicative group of regular elements in $\mathfrak{R}$. By an *elementary $\mathfrak{G}$-matrix* or simply an *elementary matrix* if no confusion is to be feared, we shall mean a square matrix of any degree n, which can be transformed into the unit matrix $\mathbf{e}_n$ by a finite sequence of elementary transformations, each of which consists either of

1. *multiplying each element in a row, from the left, or a column, from the right, by the same (arbitrary) element of $\mathfrak{G}$, or*
2. *adding a left multiple of any row to some other row, or of*
3. *adding a right multiple of any column to some other column, the multiplier in each of the last two cases being an arbitrary element of $\mathfrak{R}$.*

The inverse of each of these elementary transformations is clearly a transformation of the same type. Therefore the elementary matrices of a given degree constitute a sub-group of the multiplicative group of regular matrices of that degree. As with unimodular matrices of integers it may be verified that, if $\mathbf{a}$ is an elementary matrix of degree n and $\mathbf{r}$ is any matrix with n rows (columns), then $\mathbf{r}$ can be transformed into $\mathbf{a}\mathbf{r}(\mathbf{r}\mathbf{a})$ by a sequence of the elementary transformations 1 and 2 (1 and 3). Notice also that, if $\| m_{ij} \|$ is any square matrix of integers, whose determinant is unity, and if $g \in \mathfrak{G}$, then $\| m_{ij}\, g \|$ is an elementary matrix. If $\mathfrak{G}$ contains the regular element $-e$ the same applies to $\| m_{ij}\, g \|$ if $| m_{ij} | = -1$. In any case

$$\left\| \begin{matrix} 0 & -\mathbf{e}_n \\ \mathbf{e}_n & 0 \end{matrix} \right\|$$

is an elementary $\mathfrak{G}$-matrix of degree $2n$, since, replacing e by 1, its determinant is $(-1)^{n^2+n} = 1$.

LEMMA 2. *If* $\mathbf{a}$ *is a regular matrix of degree* n, *then*

$$\mathbf{r} = \left\| \begin{matrix} \mathbf{a} & 0 \\ \mathbf{b} & \mathbf{a}^{-1} \end{matrix} \right\|$$

is an elementary matrix, where $\mathbf{b}$ *is any square matrix of degree* n.

By reiterating transformations of the form $\rho_i \to \rho_i + \lambda\rho_j$, where ρ_i and ρ_j $(i \neq j)$ stand for rows and $\lambda \in \mathfrak{R}$, we can transform $\mathbf{r}$ into

$$\mathbf{r}' = \left\| \begin{matrix} \mathbf{a} & 0 \\ \mathbf{b} + \mathbf{c}\mathbf{a} & \mathbf{a}^{-1} \end{matrix} \right\|,$$

where $\mathbf{c}$ is any square matrix of degree n. Taking $\mathbf{c} = (\mathbf{e}_n - \mathbf{b})\mathbf{a}^{-1}$ we have

$$\mathbf{r}' = \begin{Vmatrix} \mathbf{a} & 0 \\ \mathbf{e}_n & \mathbf{a}^{-1} \end{Vmatrix},$$

and $\mathbf{r}'$ is transformable into the elementary matrix

$$\begin{Vmatrix} 0 & -\mathbf{e}_n \\ \mathbf{e}_n & 0 \end{Vmatrix}$$

by the sequence of transformations

$$\begin{Vmatrix} \mathbf{a} & 0 \\ \mathbf{e}_n & \mathbf{a}^{-1} \end{Vmatrix} \to \begin{Vmatrix} 0 & -\mathbf{e}_n \\ \mathbf{e}_n & \mathbf{a}^{-1} \end{Vmatrix} \to \begin{Vmatrix} 0 & -\mathbf{e}_n \\ \mathbf{e}_n & 0 \end{Vmatrix},$$

in which, first, the bottom rectangle $\| \mathbf{e}_n, \mathbf{a}^{-1} \|$ is multiplied on the left by $-\mathbf{a}$ and added to the top, and then $\| 0, -\mathbf{e}_n \|$ is multiplied by $\mathbf{a}^{-1}$ and added to the bottom. The lemma now follows from the fact that a matrix which is transformable by the elementary transformations into an elementary matrix is itself elementary.

Let us introduce the convention that a matrix may have m rows (columns) and no columns (rows), for any $m = 0, 1, \cdots$. If a matrix $\mathbf{a}$ has m rows and p columns, and if $\mathbf{b}$ has n rows and q columns, where $m, n, p, q \geqq 0$, then, as usual, the product $\mathbf{ab}$, with $m \geqq 0$ rows and $q \geqq 0$ columns, shall exist if, and only if, $p = n$. If $m, q > 0$, $p = n = 0$, then $\mathbf{ab}$ shall be the zero matrix with m rows and q columns. In formulae which involve $\mathbf{e}_k$, with $k \geqq 0$, $\mathbf{e}_0$ is to be the empty matrix with no rows or columns, which we include among the elementary and regular matrices. With these conventions let $\mathbf{r}^n, \cdots, \mathbf{r}^1$ be any set of matrices, such that $\mathbf{r}^p$ has α_p rows and α_{p-1} columns ($p = 1, \cdots, n$; $\alpha_p \geqq 0$). We proceed to define two kinds of equivalence between $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and a set of matrices $\mathbf{s}^n, \cdots, \mathbf{s}^1$, which satisfy a similar condition. That is to say, we can form the product $\mathbf{s}^{p+1}\mathbf{s}^p$ for each $p = 1, \cdots, n - 1$, though $\mathbf{s}^p$ need not have the same shape as $\mathbf{r}^p$. We first of all allow ourselves to border $\mathbf{r}^p$ with a new last row and column, which have the common element e in the bottom right hand corner, at the same time adding a final row of zeros to $\mathbf{r}^{p-1}$, if $p > 1$, or an empty row if $\alpha_{p-2} = 0$, and a final column of zeros to $\mathbf{r}^{p+1}$, if $p < n$, or an empty column if $\alpha_{p+1} = 0$ (if $\alpha_p = \alpha_{p-1} = 0$, then $\mathbf{r}^p$ will be transformed into the matrix with the single element e). We shall call this operation an *elementary expansion* of the matrices $\mathbf{r}^n, \cdots, \mathbf{r}^1$. We shall also call it a *primary expansion* of $\mathbf{r}^p$, and shall describe the corresponding expansions of $\mathbf{r}^{p\pm 1}$, as *secondary expansions*. In order to avoid the qualifications "if $p > 1$" and "if $p < n$" we shall assume that $\mathbf{r}^{n+1}$ and $\mathbf{r}^0$ exist, where $\mathbf{r}^{n+1}$ has no rows and α_n columns and $\mathbf{r}^0$ has α_0 rows and no columns. A primary expansion shall never be applied to $\mathbf{r}^0$, and shall only be applied to $\mathbf{r}^{n+1}$ when special permission is granted. The inverse of an elementary expansion will be called an *elementary contraction.* Two sets of

matrices $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$, such that the products $\mathbf{r}^{p+1}\mathbf{r}^p$ and $\mathbf{s}^{p+1}\mathbf{s}^p$ exist for each $p = 1, \cdots, n-1$, will be described as *L-equivalent* if and only if

$$\bar{\mathbf{s}}^p = \mathbf{a}_p \bar{\mathbf{r}}^p \mathbf{a}_{p-1}^{-1} \qquad (p = 1, \cdots, n), \tag{2.1}$$

where $\bar{\mathbf{r}}^n, \cdots, \bar{\mathbf{r}}^1$ and $\bar{\mathbf{s}}^n, \cdots, \bar{\mathbf{s}}^1$ are obtained from $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ by sequences of elementary expansions, and $\mathbf{a}_p$ is an elementary matrix of the appropriate degree, for each $p = 0, \cdots, n$. Two sets of matrices, $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$, will be described as L^*-equivalent if, and only if, they satisfy the same condition, except that $\mathbf{a}_p$ may be any regular matrix of the appropriate degree. These relations are obviously symmetric between the two sets of matrices.

We now assume that the group $\mathfrak{G}$ contains $-e$, in which case two rows or columns of a given matrix may be interchanged by a sequence of elementary transformations. This being so, if we apply a sequence of transformations to the matrices $\mathbf{r}^n, \cdots, \mathbf{r}^1$, which consist of elementary expansions and contractions, and transformations of the form (2.1), a simple induction shows that all the expansions may be applied first,[7] then a transformation of the form (2.1) and finally the contractions. Hence it follows that the relations of L-equivalence and L^*-equivalence are equivalence relations in the technical sense, meaning that they are symmetric and transitive.

LEMMA 3. *If* $\mathbf{r}^n, \cdots, \mathbf{r}^1$ *and* $\mathbf{s}^n, \cdots, \mathbf{s}^1$ *are L^*-equivalent, then they can be expanded into sets of matrices* $\bar{\mathbf{r}}^n, \cdots, \bar{\mathbf{r}}^1$ *and* $\bar{\mathbf{s}}^n, \cdots, \bar{\mathbf{s}}^1$, *which are related by equations of the form* (2.1), *subject to the condition that* $\mathbf{a}_0, \cdots, \mathbf{a}_{n-1}$ *shall be elementary matrices. Thus the sets* $\mathbf{a}\bar{\mathbf{r}}^n, \bar{\mathbf{r}}^{n-1}, \cdots, \bar{\mathbf{r}}^1$ *and* $\bar{\mathbf{s}}^n, \cdots, \bar{\mathbf{s}}^1$ *are L-equivalent, where* $\mathbf{a}$ *is some regular matrix.*

After initial expansions we assume that $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ are themselves related by transformations of the form (2.1), in which $\mathbf{a}_0, \cdots, \mathbf{a}_{p-1}$ are elementary matrices ($0 \leqq p \leqq n$), this last condition being vacuous if $p = 0$. If $p = n$ there is nothing to prove. Otherwise we expand $\mathbf{r}^{p+2}, \mathbf{r}^{p+1}, \mathbf{r}^p$ into

$$\bar{\mathbf{r}}^{p+2} = \| \mathbf{r}^{p+2}, 0 \|, \qquad \bar{\mathbf{r}}^{p+1} = \begin{Vmatrix} \mathbf{r}^{p+1} & 0 \\ 0 & \mathbf{e}_{\alpha_p} \end{Vmatrix}, \qquad \bar{\mathbf{r}}^p = \begin{Vmatrix} \mathbf{r}^p \\ 0 \end{Vmatrix},$$

and write

$$\bar{\mathbf{s}}^{p+2} = \mathbf{a}_{p+2} \| \mathbf{r}^{p+2}, 0 \| \begin{Vmatrix} \mathbf{a}_{p+1}^{-1} & 0 \\ 0 & \mathbf{a}_p \end{Vmatrix} = \| \mathbf{s}^{p+2}, 0 \|$$

$$\bar{\mathbf{s}}^{p+1} = \begin{Vmatrix} \mathbf{a}_{p+1} & 0 \\ 0 & \mathbf{a}_p^{-1} \end{Vmatrix} \begin{Vmatrix} \mathbf{r}^{p+1} & 0 \\ 0 & \mathbf{e}_{\alpha_p} \end{Vmatrix} \begin{Vmatrix} \mathbf{a}_p^{-1} & 0 \\ 0 & \mathbf{a}_p \end{Vmatrix} = \begin{Vmatrix} \mathbf{s}^{p+1} & 0 \\ 0 & \mathbf{e}_{\alpha_p} \end{Vmatrix},$$

$$\bar{\mathbf{s}}^p = \begin{Vmatrix} \mathbf{a}_p & 0 \\ 0 & \mathbf{a}_p^{-1} \end{Vmatrix} \begin{Vmatrix} \mathbf{r}^p \\ 0 \end{Vmatrix} \mathbf{a}_{p-1}^{-1} = \begin{Vmatrix} \mathbf{s}^p \\ 0 \end{Vmatrix}.$$

The lemma now follows from lemma 2 and induction on p.

[7] Cf. the preliminary argument in theorem 1 below.

We now define what we shall temporarily call EL-equivalence (extended L-equivalence). The conditions for EL-equivalence are the same as for L-equivalence except that we add empty matrices $\mathbf{r}^{n+1}, \mathbf{r}^{n+2}, \cdots$ to a given set $\mathbf{r}^n, \cdots, \mathbf{r}^1$, and allow a primary expansion of $\mathbf{r}^p$, even if $p > n$, where initially $\mathbf{r}^p$ has no rows or columns if $p > n + 1$, and has no rows and α_n columns if $p = n + 1$. If $p > n$ a primary expansion of $\mathbf{r}^p$ will be called an *extended expansion* of the set $\mathbf{r}^n, \cdots, \mathbf{r}^1$. It is clear that EL-equivalence, like L-equivalence and for a similar reason, is an equivalence relation in the technical sense.

THEOREM 1. *If the matrices* $\mathbf{r}^n, \cdots, \mathbf{r}^1$ *and* $\mathbf{s}^n, \cdots, \mathbf{s}^1$ *are EL-equivalent, then they are L-equivalent.*

To say that $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ are EL-equivalent is to say that matrices $\bar{\mathbf{r}}^m, \cdots, \bar{\mathbf{r}}^1$ and $\bar{\mathbf{s}}^m, \cdots, \bar{\mathbf{s}}^1$ are related by a transformation of the form (2.1), with n replaced by $m \geqq n$, where $\bar{\mathbf{r}}^m, \cdots, \bar{\mathbf{r}}^1$ and $\bar{\mathbf{s}}^m, \cdots, \bar{\mathbf{s}}^1$ are obtained from $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ by expansions and extended expansions and $\mathbf{a}_0, \cdots, \mathbf{a}_m$ are elementary matrices. The theorem will follow from induction on m when we have proved that $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ are EL-equivalent under a transformation in which no primary expansion is applied to $\mathbf{r}^p$ if $p > m - 1$, unless $m = n$, in which case there is nothing to prove. I say that, in the given sequence of expansions, the primary expansions of $\mathbf{r}^m$, if there are any, may be applied last of all. For if $p < m - 1$ a primary expansion of $\mathbf{r}^p$ is obviously interchangeable with a primary expansion of $\mathbf{r}^m$. The effect of expanding $\mathbf{r}^{m-1}$ before, instead of after $\mathbf{r}^m$, is to interchange the rows which are added by the given secondary and primary expansions of $\mathbf{r}^{m-1}$, and also the corresponding columns of $\mathbf{r}^m$. Therefore we may postpone the primary expansions of $\mathbf{r}^m$ till the final stages of the expansion and, if necessary, include certain permutations of the rows of $\bar{\mathbf{r}}^{m-1}$ and the columns of $\bar{\mathbf{r}}^m$ in the transformation (2.1). So we have to prove that matrices $\mathbf{r}'^{m-1}, \cdots, \bar{\mathbf{r}}^1$ and $\mathbf{s}'^{m-1}, \cdots, \bar{\mathbf{s}}^1$ are L-equivalent if $\bar{\mathbf{r}}^m, \bar{\mathbf{r}}^{m-1}, \bar{\mathbf{r}}^{m-2}, \cdots, \bar{\mathbf{r}}^1$ and $\bar{\mathbf{s}}^m, \bar{\mathbf{s}}^{m-1}, \bar{\mathbf{s}}^{m-2}, \cdots, \bar{\mathbf{s}}^1$ are related by a transformation of the form (2.1), in which $\mathbf{a}_0, \cdots, \mathbf{a}_m$ are elementary matrices, where

$$
\begin{aligned}
&\bar{\mathbf{r}}^m = \|0, \mathbf{e}_k\|, \qquad \bar{\mathbf{r}}^{m-1} = \left\| \begin{matrix} \mathbf{r}'^{m-1} \\ 0 \end{matrix} \right\|, \\
&\bar{\mathbf{s}}^m = \|0, \mathbf{e}_l\|, \qquad \bar{\mathbf{s}}^{m-1} = \left\| \begin{matrix} \mathbf{s}'^{m-1} \\ 0 \end{matrix} \right\|,
\end{aligned}
\tag{2.2}
$$

and $k = l$, since the matrices $\bar{\mathbf{r}}^m$ and $\bar{\mathbf{s}}^m$ have the same shape. This being so, we have

$$
\begin{aligned}
\mathbf{a}\,\|0, \mathbf{e}_k\| \left\| \begin{matrix} \mathbf{p}_1 & \mathbf{q}_1 \\ \mathbf{p}_2 & \mathbf{q}_2 \end{matrix} \right\| &= \|0, \mathbf{a}\| \left\| \begin{matrix} \mathbf{p}_1 & \mathbf{q}_1 \\ \mathbf{p}_2 & \mathbf{q}_2 \end{matrix} \right\| \\
&= \|\mathbf{a}\mathbf{p}_2, \mathbf{a}\mathbf{q}_2\| \\
&= \|0, \mathbf{e}_k\|,
\end{aligned}
$$

where $\mathbf{a} = \mathbf{a}_m$ and

$$\mathbf{a}_{m-1}^{-1} = \begin{Vmatrix} \mathbf{p}_1 & \mathbf{q}_1 \\ \mathbf{p}_2 & \mathbf{q}_2 \end{Vmatrix},$$

$\mathbf{q}_2$ being a square matrix of degree k.

Therefore $\mathbf{p}_2 = 0$ and $\mathbf{q}_2 = \mathbf{a}^{-1}$. Since $\mathbf{a}^{-1}$ is regular it follows from lemma 1 that $\mathbf{a}_{m-1}$ is of the form

$$\mathbf{a}_{m-1} = \begin{Vmatrix} \mathbf{b} & \mathbf{c} \\ 0 & \mathbf{a} \end{Vmatrix},$$

whence

$$\mathbf{a}_{m-1}\bar{\mathbf{r}}_{m-1} = \begin{Vmatrix} \mathbf{b}\mathbf{r}'^{m-1} \\ 0 \end{Vmatrix}.$$

It follows from an argument in the proof of lemma 2 that

$$\begin{Vmatrix} \mathbf{b} & \mathbf{c} - \mathbf{c}\mathbf{a}^{-1}\mathbf{a} \\ 0 & \mathbf{a} \end{Vmatrix} = \begin{Vmatrix} \mathbf{b} & 0 \\ 0 & \mathbf{a} \end{Vmatrix}$$

is an elementary matrix. Therefore, writing $n = m - 1$ and omitting the primes and dashes, the theorem will follow when we have proved that two sets of matrices $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{b}\mathbf{r}^n, \cdots, \mathbf{r}^1$ are L-equivalent if $\mathbf{b}$ is a square matrix, such that

$$\begin{Vmatrix} \mathbf{b} & 0 \\ 0 & \mathbf{a} \end{Vmatrix}$$

is an elementary matrix, where $\mathbf{a}$ is some elementary matrix.[8] To prove this we first expand $\mathbf{r}^n$, transforming $\mathbf{r}^n$ and $\mathbf{r}^{n-1}$ into

$$\begin{Vmatrix} \mathbf{r}^n & 0 \\ 0 & \mathbf{e}_k \end{Vmatrix} \quad \text{and} \quad \begin{Vmatrix} \mathbf{r}^{n-1} \\ 0 \end{Vmatrix},$$

where k is the degree of $\mathbf{a}$. Since $\mathbf{a}$ is an elementary matrix, so obviously is

$$\begin{Vmatrix} \mathbf{e}_\alpha & 0 \\ 0 & \mathbf{a} \end{Vmatrix} \qquad (\alpha = \alpha_n).$$

[8] See note II_1 in §6 below.

Therefore $\mathbf{r}^n, \cdots, \mathbf{r}^1$ are L-equivalent to $\bar{\mathbf{r}}^n, \bar{\mathbf{r}}^{n-1}, \cdots, \mathbf{r}^1$, where

$$\bar{\mathbf{r}}^n = \left\| \begin{matrix} \mathbf{b} & 0 \\ 0 & \mathbf{a} \end{matrix} \right\| \left\| \begin{matrix} \mathbf{r}^n & 0 \\ 0 & \mathbf{e}_k \end{matrix} \right\| \left\| \begin{matrix} \mathbf{e}_\alpha & 0 \\ 0 & \mathbf{a}^{-1} \end{matrix} \right\|$$

$$= \left\| \begin{matrix} \mathbf{br}^n & 0 \\ 0 & \mathbf{a} \end{matrix} \right\| \left\| \begin{matrix} \mathbf{e}_\alpha & 0 \\ 0 & \mathbf{a}^{-1} \end{matrix} \right\|$$

$$= \left\| \begin{matrix} \mathbf{br}^n & 0 \\ 0 & \mathbf{e}_k \end{matrix} \right\|,$$

$$\bar{\mathbf{r}}^{n-1} = \left\| \begin{matrix} \mathbf{e}_\alpha & 0 \\ 0 & \mathbf{a} \end{matrix} \right\| \left\| \begin{matrix} \mathbf{r}^{n-1} \\ 0 \end{matrix} \right\|$$

$$= \left\| \begin{matrix} \mathbf{r}^{n-1} \\ 0 \end{matrix} \right\|.$$

But $\bar{\mathbf{r}}^n, \bar{\mathbf{r}}^{n-1}$ contract into $\mathbf{br}^n, \mathbf{r}^{n-1}$, and the proof is complete.

Two sets of matrices, $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$, will be described as *$L^m$-equivalent,* for any $m \geqq 0$; if, and only if, they satisfy the conditions for EL-equivalence, except that, in addition to expansions and extended expansions, we may add a final row of zeros to $\mathbf{r}^k$ and a final column of zeros to $\mathbf{r}^{k+1}$, for any $k > m$, on the understanding that, if $\alpha_{k-1} = 0$, then $\left\| \begin{matrix} \mathbf{r}^k \\ 0 \end{matrix} \right\|$ is an empty matrix having $\alpha_k + 1$ rows and no columns, with a similar convention concerning $\| \mathbf{r}^{k+1}, 0 \|$ if $\alpha_{k+1} = 0$. Thus, in passing from $\mathbf{r}^n, \cdots, \mathbf{r}^1$ to $\mathbf{s}^n, \cdots, \mathbf{s}^1$ by a series of elementary transformations, one may, at any stage, apply such a transformation or its inverse. Notice that L-equivalence implies L^m-equivalence for every value of m, and that L^m-equivalence implies L^q-equivalence if $m > q$. Clearly L^m-equivalence is also a symmetric and transitive relation. We shall prove that, if $m \geqq n$, then L^m-equivalence is the same as L^*-equivalence, subject to the following condition on the ring $\mathfrak{R}$. If

$$\text{(2.3)} \qquad \mathbf{a} \left\| \begin{matrix} 0 & \mathbf{e}_k \\ 0 & 0 \end{matrix} \right\| = \left\| \begin{matrix} 0 & \mathbf{e}_l \\ 0 & 0 \end{matrix} \right\| \mathbf{b},$$

where $\mathbf{a}$ and $\mathbf{b}$ are regular matrices of appropriate degrees, then $k = l$. This condition is satisfied[9] if there is a homomorphism, $\phi(\mathfrak{R}) = \mathfrak{R}_0$, of $\mathfrak{R}$ on a commutative ring $\mathfrak{R}_0$, which contains at least two elements. For the determinant of any square matrix, whose elements are in $\mathfrak{R}_0$, can be calculated in the ordinary way, and hence the rank of any matrix. If $\mathbf{s} = \mathbf{arb}$, $\mathbf{r} = \mathbf{a}'\mathbf{sb}'$, where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{a}'$, $\mathbf{b}'$, $\mathbf{r}$ and $\mathbf{s}$ are matrices with elements in $\mathfrak{R}_0$, it follows from the standard argument that $\mathbf{r}$ and $\mathbf{s}$ have the same rank. Since $\mathfrak{R}_0$ contains at least two ele-

---

[9] See note II_2 in §6 below.

ments and since $\phi(e)\phi(x) = \phi(ex) = \phi(x)$ for any $\phi(x) \in \mathfrak{R}_0$, it follows that $\mathfrak{R}_0$ has a unit element, namely $\phi(e)$. Therefore $\phi(e)$ is not nilpotent, whence a determinant with $\phi(e)$ down the main diagonal, and zeros everywhere else, is not zero. On replacing e, a_{ij} and $b_{\lambda\mu}$, in (2.3), by $\phi(e)$, $\phi(a_{ij})$ and $\phi(b_{\lambda\mu})$, where $\mathbf{a} = \| a_{ij} \|$, $\mathbf{b} = \| b_{\lambda\mu} \|$, it follows from equality of rank that $k = l$.

THEOREM 2. *Subject to the above condition on $\mathfrak{R}$, if two sets of matrices $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ are L^n-equivalent, then they are L^*-equivalent. Conversely, L^*-equivalence implies L^m-equivalence for every value of m.*

Let $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ be L^n-equivalent. Then $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ can be expanded, by elementary and extended expansions and transformations of the form $\mathbf{r}^{k+1}, \mathbf{r}^k \to \| \mathbf{r}^{k+1}, 0 \|, \left\| \begin{matrix} \mathbf{r}^k \\ 0 \end{matrix} \right\|$ $(k > n)$, into sets $\bar{\mathbf{r}}^q, \cdots, \bar{\mathbf{r}}^1$ and $\bar{\mathbf{s}}^q, \cdots, \bar{\mathbf{s}}^1$ $(q \geqq n)$, such that $\bar{\mathbf{s}}^p = \mathbf{a}_p \bar{\mathbf{r}}^p \mathbf{a}_{p-1}^{-1}$ $(p = 1, \cdots, q)$, where $\mathbf{a}_0, \cdots, \mathbf{a}_q$ are elementary matrices. As in the proof of theorem 1 we may suppose that the primary expansions of $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ are applied first in the expansions $\mathbf{r}^p \to \bar{\mathbf{r}}^p$ and $\mathbf{s}^p \to \bar{\mathbf{s}}^p$. Without altering the notation, let us assume that they have already been applied. Then, after rearranging the rows and columns of $\bar{\mathbf{r}}^{n+1}$ and $\bar{\mathbf{s}}^{n+1}$ if necessary, we have

$$\bar{\mathbf{r}}^{n+1} = \left\| \begin{matrix} 0 & \mathbf{e}_k \\ 0 & 0 \end{matrix} \right\|, \qquad \bar{\mathbf{r}}^n = \left\| \begin{matrix} \mathbf{r}^n \\ 0 \end{matrix} \right\|$$

$$\bar{\mathbf{s}}^{n+1} = \left\| \begin{matrix} 0 & \mathbf{e}_l \\ 0 & 0 \end{matrix} \right\|, \qquad \bar{\mathbf{s}}^n = \left\| \begin{matrix} \mathbf{s}^n \\ 0 \end{matrix} \right\| \qquad (k, l \geqq 0),$$

where the bottom rectangles of zeros in $\bar{\mathbf{r}}^n$ and $\bar{\mathbf{s}}^n$ contain k and l rows respectively, and the bottom rectangles of zeros in $\bar{\mathbf{r}}^{n+1}$ and $\bar{\mathbf{s}}^{n+1}$ may be empty. If $k = l = 0$ the matrices $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ are L-equivalent. Otherwise it follows from the relation $\mathbf{a}_{n+1}\bar{\mathbf{r}}^{n+1} = \bar{\mathbf{s}}^{n+1}\mathbf{a}_n$ and the condition (2.3) that $k = l$, and we have

$$\left\| \begin{matrix} \mathbf{b}_1 & \mathbf{c}_1 \\ \mathbf{b}_2 & \mathbf{c}_2 \end{matrix} \right\| \left\| \begin{matrix} 0 & \mathbf{e}_k \\ 0 & 0 \end{matrix} \right\| = \left\| \begin{matrix} 0 & \mathbf{e}_k \\ 0 & 0 \end{matrix} \right\| \left\| \begin{matrix} \mathbf{p}_1 & \mathbf{q}_1 \\ \mathbf{p}_2 & \mathbf{q}_2 \end{matrix} \right\|, \tag{2.4}$$

where

$$\left\| \begin{matrix} \mathbf{b}_1 & \mathbf{c}_1 \\ \mathbf{b}_2 & \mathbf{c}_2 \end{matrix} \right\| = \mathbf{a}_{n+1}, \qquad \left\| \begin{matrix} \mathbf{p}_1 & \mathbf{q}_1 \\ \mathbf{p}_2 & \mathbf{q}_2 \end{matrix} \right\| = \mathbf{a}_n,$$

$\mathbf{b}_1$ and $\mathbf{q}_2$ being square matrices of degree k. Also

$$\left\| \begin{matrix} \mathbf{s}^n \\ 0 \end{matrix} \right\| = \left\| \begin{matrix} \mathbf{p}_1 & \mathbf{q}_1 \\ \mathbf{p}_2 & \mathbf{q}_2 \end{matrix} \right\| \left\| \begin{matrix} \mathbf{r}^n \\ 0 \end{matrix} \right\| \mathbf{a}_{n-1}^{-1} = \left\| \begin{matrix} \mathbf{p}_1\mathbf{r}^n\mathbf{a}_{n-1}^{-1} \\ \mathbf{p}_2\mathbf{r}^n\mathbf{a}_{n-1}^{-1} \end{matrix} \right\|.$$

Since $k = l$ it follows that $\mathbf{s}^n = \mathbf{p}_1\mathbf{r}^n\mathbf{a}_{n-1}^{-1}$, and we have only to show that the matrix $\mathbf{p}_1$ is regular. From (2.4) we have

$$\begin{Vmatrix} 0 & \mathbf{b}_1 \\ 0 & \mathbf{b}_2 \end{Vmatrix} = \begin{Vmatrix} \mathbf{p}_2 & \mathbf{q}_2 \\ 0 & 0 \end{Vmatrix},$$

whence $\mathbf{b}_2 = 0$, $\mathbf{p}_2 = 0$, $\mathbf{b}_1 = \mathbf{q}_2$. Let

$$\begin{Vmatrix} \mathbf{b}_1 & \mathbf{c}_1 \\ \mathbf{b}_2 & \mathbf{c}_2 \end{Vmatrix}^{-1} = \begin{Vmatrix} \mathbf{b}_1' & \mathbf{b}_2' \\ \mathbf{c}_1' & \mathbf{c}_2' \end{Vmatrix}.$$

Since $\mathbf{b}_2 = 0$, we have

$$\begin{Vmatrix} \mathbf{b}_1' & \mathbf{b}_2' \\ \mathbf{c}_1' & \mathbf{c}_2' \end{Vmatrix} \begin{Vmatrix} \mathbf{b}_1 & \mathbf{c}_1 \\ \mathbf{b}_2 & \mathbf{c}_2 \end{Vmatrix} = \begin{Vmatrix} \mathbf{b}_1'\mathbf{b}_1 & * \\ * & * \end{Vmatrix} = \mathbf{e}_r$$

for some $r > 0$. Therefore $\mathbf{b}_1$ has a left inverse. Since $\mathbf{p}_2 = 0$, $\mathbf{q}_2 = \mathbf{b}_1$, and since

$$\begin{Vmatrix} \mathbf{p}_1 & \mathbf{q}_1 \\ 0 & \mathbf{q}_2 \end{Vmatrix}$$

is a regular matrix, it follows from lemma 1 that $\mathbf{p}_1$ is a regular matrix, and the first part of the theorem is established.

Conversely, let $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ be L^*-equivalent and, given $m \geqq 0$, let us introduce empty matrices $\mathbf{r}^{m+1}, \cdots, \mathbf{r}^{n+1}$ and $\mathbf{s}^{m+1}, \cdots, \mathbf{s}^{n+1}$ if $m \geqq n$. Then the sets $\mathbf{r}^k, \cdots, \mathbf{r}^1$ and $\mathbf{s}^k, \cdots, \mathbf{s}^1$ are obviously L^*-equivalent, where $k = \max(m, n) + 1$, and by lemma 3 there are sets $\bar{\mathbf{r}}^k, \cdots, \bar{\mathbf{r}}^1$ and $\bar{\mathbf{s}}^k, \cdots, \bar{\mathbf{s}}^1$, which are L-equivalent to $\mathbf{r}^k, \cdots, \mathbf{r}^1$ and $\mathbf{s}^k, \cdots, \mathbf{s}^1$ respectively, such that $\bar{\mathbf{s}}^k = \mathbf{a}\bar{\mathbf{r}}^k$, $\bar{\mathbf{s}}^p = \bar{\mathbf{r}}^p$ for $p = 1, \cdots, k-1$, where $\mathbf{a}$ is a regular matrix. Let

$$\bar{\bar{\mathbf{r}}}^k = \begin{Vmatrix} \bar{\mathbf{r}}^k \\ 0 \end{Vmatrix}, \qquad \bar{\bar{\mathbf{s}}}^k = \begin{Vmatrix} \bar{\mathbf{s}}^k \\ 0 \end{Vmatrix},$$

where the number of zero rows indicated by 0 is the same as the number of rows in $\bar{\mathbf{r}}^k$ and $\bar{\mathbf{s}}^k$. Then

$$\bar{\bar{\mathbf{s}}}^k = \begin{Vmatrix} \mathbf{a} & 0 \\ 0 & \mathbf{a}^{-1} \end{Vmatrix} \bar{\bar{\mathbf{r}}}^k,$$

and the theorem follows from lemma 2.

Corollary. *If* $\mathbf{r}^n, \cdots, \mathbf{r}^1$ *and* $\mathbf{s}^n, \cdots, \mathbf{s}^1$ *are* L^n*-equivalent, then they are* L^m*-equivalent for every* $m \geqq 0$.

In the second part of theorem 2 it follows from the argument in theorem 1, which is referred to in the first part of theorem 2, that the transformations $\mathbf{r}^k$, $\mathbf{s}^k \to \begin{Vmatrix} \mathbf{r}^k \\ 0 \end{Vmatrix}, \begin{Vmatrix} \mathbf{s}^k \\ 0 \end{Vmatrix}$ may be applied before any primary expansions of $\mathbf{r}^k, \cdots, \mathbf{r}^1$ an $\mathbf{s}^k, \cdots, \mathbf{s}^1$. Therefore, taking $m \geqq n$, we have the addendum to theorem 2:

ADDENDUM. *If* $\mathbf{r}^n, \cdots, \mathbf{r}^1$ *and* $\mathbf{s}^n, \cdots, \mathbf{s}^1$ *are* L^**-equivalent, then* $\mathbf{r}^{m+1}, \cdots, \mathbf{r}^1$ *and* $\mathbf{s}^{m+1}, \cdots, \mathbf{s}^1$ *are L-equivalent, for any* $m \geqq n$, *where* $\mathbf{r}^{m+1}$ *and* $\mathbf{s}^{m+1}$ *are certain empty or zero matrices if* $m = n$, *and* $\mathbf{r}^p$ *and* $\mathbf{s}^p$ $(p = n + 1, \cdots, m + 1)$ *are certain empty matrices if* $m > n$ $(\alpha_{m+1} \geqq 0, \alpha_m = \cdots = \alpha_{n+1} = 0)$.

We now assume that $\mathbf{r}^p\mathbf{r}^{p-1} = 0$ for each $p = 2, \cdots, n$, where 0 stands for an empty matrix if either $\alpha_p = 0$ or $\alpha_{p-2} = 0$. This condition is obviously invariant under all our elementary transformations and under transformations of the form (2.1). On this assumption, we shall prove a lemma, which brings the elementary transformations by which L-equivalence is defined into closer relation with elementary deformations of complexes. Let us extend the terms *elementary expansion* and *primary expansion* so as to include any transformation of the form

$$\begin{aligned} \mathbf{r}^{p+1} &\to \left\| \mathbf{r}^{p+1}, \ 0 \right\| \\ \mathbf{r}^{p} &\to \left\| \begin{matrix} \mathbf{r}^p & 0 \\ \mathbf{a} & g \end{matrix} \right\| = \left\| \begin{matrix} \mathbf{r}^p & 0 \\ 0 & e \end{matrix} \right\| \left\| \begin{matrix} \mathbf{e}_\alpha & 0 \\ \mathbf{a} & g \end{matrix} \right\| \\ \mathbf{r}^{p-1} &\to \left\| \begin{matrix} \mathbf{r}^{p-1} \\ \mathbf{b} \end{matrix} \right\| = \left\| \begin{matrix} \mathbf{e}_\alpha & 0 \\ -g^{-1}\mathbf{a} & g^{-1} \end{matrix} \right\| \left\| \begin{matrix} \mathbf{r}^{p-1} \\ 0 \end{matrix} \right\|, \end{aligned} \tag{2.5} \qquad (\alpha = \alpha_{p-1})$$

where $\mathbf{a}$ stands for any row $\| a_1, \cdots, a_\alpha \|$ $(a_\lambda \in \mathfrak{R})$, $g \in \mathfrak{G}$ and $\mathbf{b}$ stands for the row $-g^{-1}\mathbf{a}\mathbf{r}^{p-1}$, of which the j^{th} element is

$$-g^{-1} \sum_{\lambda=1}^{\alpha} a_\lambda r_{\lambda j}^{p-1} \qquad (j = 1, \cdots, \alpha_{p-2}; \mathbf{r}^{p-1} = \| \mathbf{r}_{\lambda j}^{p-1} \|).$$

The inverse of such a transformation will be called an *elementary contraction.* In this definition it is to be understood that the new rows and columns are indicated as the last rows and columns in (2.5) merely for convenience of notation. Thus the removal of the μ^{th} column and the k^{th} row of $\mathbf{r}^p$ in case $\mathbf{r}_{k\mu}^p \in G$, $r_{j\mu}^p = 0$ if $j \neq k$, together with the μ^{th} row of $\mathbf{r}^{p-1}$ and the k^{th} column of $\mathbf{r}^{p+1}$ will be called an elementary contraction. Notice that, if $r_{k\mu}^p = g$, $r_{j\mu}^p = 0$ $(j \neq k)$, where $g \in \mathfrak{G}$, or if g is any regular element, then the μ^{th} row of $\mathbf{r}^{p-1}$ has the form indicated above in consequence of $\mathbf{r}^p\mathbf{r}^{p-1} = 0$ and the k^{th} column of $\mathbf{r}^{p+1}$ consists entirely of zeros since $\mathbf{r}^{p+1}\mathbf{r}^p = 0$.

Let $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ be L-equivalent sets of matrices $(\mathbf{r}^p\mathbf{r}^{p-1} = 0$, $\mathbf{s}^p\mathbf{s}^{p-1} = 0)$. We add empty matrices $\mathbf{r}^{n+1}$ and $\mathbf{s}^{n+1}$ to these sets, on the understanding that a primary expansion of $\mathbf{r}^{n+1}$ is to be regarded as an elementary expansion, not as an extended expansion.

LEMMA 4. *The matrices* $\mathbf{r}^{n+1}, \cdots, \mathbf{r}^1$ *can be transformed into* $\mathbf{s}^{n+1}, \cdots, \mathbf{s}^1$ *by a sequence of elementary expansions and contractions.*

First let $\mathbf{r}^p = \mathbf{s}^p$ for $p = 1, \cdots, n - 1$ and let $\mathbf{s}^n = \mathbf{a}\mathbf{r}^n$, where $\mathbf{a}$ is an elementary matrix. Then $\mathbf{r}^n$ can be transformed into $\mathbf{s}^n$ by a sequence of elementary transformations of the form $\rho_i \to g\rho_i + \lambda\rho_j$ $(i \neq j)$, where $g \in G$, $\lambda \in \mathfrak{R}$ and ρ_k denotes the k^{th} row of a given matrix. Taking $i = 1, j = 2$, for sim-

plicity of notation, this transformation is the resultant of the elementary expansion and contraction

$$\mathbf{r}^{n+1} \to \|e, \quad -g, \quad -\lambda, \quad 0\| \to \mathbf{r}^{n+1},$$

$$\mathbf{r}^{n} \to \begin{Vmatrix} g\rho_1 + \lambda\rho_2 \\ \rho_1 \\ \vdots \\ \rho_{\alpha_n} \end{Vmatrix} \to \begin{Vmatrix} g\rho_1 + \lambda\rho_2 \\ \rho_2 \\ \vdots \\ \rho_{\alpha_n} \end{Vmatrix},$$

in which the row $g\rho_1 + \lambda\rho_2$ is created and ρ_1 is then deleted. Thus the special case of the lemma is established.

In general, we assume, as in lemma 3, that, after certain initial expansions, the two sets of matrices are already related by equations of the form (2.1), in which $\mathbf{a}_0, \cdots, \mathbf{a}_n$ are elementary matrices. Let $\mathbf{a}_0, \cdots, \mathbf{a}_{p-1}$ be unit matrices, for some value of p $(0 \leqq p < n)$, this condition being vacuous if $p = 0$. Then we expand $\mathbf{r}^{p+2}$, $\mathbf{r}^{p+1}$, $\mathbf{r}^{p}$ and $\mathbf{s}^{p+2}$, $\mathbf{s}^{p+1}$, $\mathbf{s}^{p}$ into the matrices $\bar{\mathbf{r}}^{q}$, $\bar{\mathbf{s}}^{q}$ $(q = p, p + 1, p + 2)$, where

$$\bar{\mathbf{r}}^{p+2} = \|0, \quad \mathbf{r}^{p+2}\|, \qquad \bar{\mathbf{r}}^{p+1} = \begin{Vmatrix} \mathbf{e}_{\alpha_p} & -\mathbf{a}_p \\ 0 & \mathbf{r}^{p+1} \end{Vmatrix}, \qquad \bar{\mathbf{r}}^{p} = \begin{Vmatrix} \mathbf{a}_p \mathbf{r}^p \\ \mathbf{r}^p \end{Vmatrix} = \begin{Vmatrix} \mathbf{s}^p \\ \mathbf{r}^p \end{Vmatrix},$$

$$\bar{\mathbf{s}}^{p+2} = \|0, \quad \mathbf{s}^{p+2}\|, \qquad \bar{\mathbf{s}}^{p+1} = \begin{Vmatrix} -\mathbf{a}_p^{-1} & \mathbf{e}_{\alpha_p} \\ \mathbf{s}^{p+1} & 0 \end{Vmatrix}, \qquad \bar{\mathbf{s}}^{p} = \begin{Vmatrix} \mathbf{s}^p \\ \mathbf{a}_p^{-1}\mathbf{s}^p \end{Vmatrix} = \begin{Vmatrix} \mathbf{s}^p \\ \mathbf{r}^p \end{Vmatrix} = \bar{\mathbf{r}}^p.$$

If $\mathbf{a}$ and $\mathbf{b}$ are elementary matrices, so obviously is $\begin{Vmatrix} \mathbf{a} & 0 \\ 0 & \mathbf{b} \end{Vmatrix}$. Therefore it follows from an argument in the proof of lemma 2 that

$$\bar{\mathbf{a}}_{p+1} = \begin{Vmatrix} -\mathbf{a}_p^{-1} & 0 \\ \mathbf{s}^{p+1} & \mathbf{a}_{p+1} \end{Vmatrix}$$

is an elementary matrix, and since $\mathbf{s}^{p+1}\mathbf{a}_p = \mathbf{a}_{p+1}\mathbf{r}^{p+1}$ we have

$$\bar{\mathbf{a}}_{p+1}\bar{\mathbf{r}}^{p+1} = \begin{Vmatrix} -\mathbf{a}_p^{-1} & \mathbf{e}_{\alpha_p} \\ \mathbf{s}^{p+1} & 0 \end{Vmatrix} = \bar{\mathbf{s}}^{p+1}.$$

Also it may be verified that

$$\bar{\mathbf{a}}_{p+1}^{-1} = \begin{Vmatrix} -\mathbf{a}_p & 0 \\ \mathbf{r}^{p+1} & \mathbf{a}_{p+1}^{-1} \end{Vmatrix},$$

and since $\mathbf{r}^{p+2}\mathbf{r}^{p+1} = 0$ it follows that

$$\begin{aligned} \mathbf{a}_{p+2}\bar{\mathbf{r}}^{p+2}\bar{\mathbf{a}}_{p+1}^{-1} &= \|0, \quad \mathbf{a}_{p+2}\mathbf{r}^{p+2}\| \begin{Vmatrix} -\mathbf{a}_p & 0 \\ \mathbf{r}^{p+1} & \mathbf{a}_{p+1}^{-1} \end{Vmatrix} \\ &= \|0, \quad \mathbf{a}_{p+2}\mathbf{r}^{p+2}\mathbf{a}_{p+1}^{-1}\| \\ &= \bar{\mathbf{s}}^{p+2}. \end{aligned}$$

The conditions are now as before, with p replaced by $p + 1$. Therefore the lemma follows from the special case and induction on p.

We conclude this section by recovering, in a modified form, a result due to Reidemeister,[10] concerning three given matrices **a**, **b** and **c**, which satisfy certain special conditions. First, the ring $\mathfrak{R}$, to which the elements of **a**, **b** and **c** belong shall be commutative. Secondly, **a**, **b** and **c** shall be of the form $\mathbf{a} = \| a_\lambda^i \|$, $\mathbf{b} = \| b_\rho^\lambda \|$ and $\mathbf{c} = \| c_t^\rho \|$ $(i = 1, \cdots, n;\ \lambda = 1, \cdots, m;\ \rho = 1, \cdots, p;\ t = 1, \cdots, q)$, where the superscripts indicate the rows, and finally we assume that $m - n = p - q = r \geqq 0$. Then, using a notation which is familiar in tensor analysis,[11] we define

$$A_{\lambda_1 \cdots \lambda_n} = e_{i_1 \cdots i_n} a_{\lambda_1}^{i_1} \cdots a_{\lambda_n}^{i_n},$$

$$B_{\lambda_1 \cdots \lambda_n}^{\rho_1 \cdots \rho_q} = \frac{1}{r!} e_{\lambda_1 \cdots \lambda_n \mu_1 \cdots \mu_r} e^{\rho_1 \cdots \rho_q \sigma_1 \cdots \sigma_r} b_{\sigma_1}^{\mu_1} \cdots b_{\sigma_r}^{\mu_r},$$

$$C^{\rho_1 \cdots \rho_q} = e^{t_1 \cdots t_q} c_{t_1}^{\rho_1} \cdots c_{t_q}^{\rho_q},$$

with the convention that $B = e$, the unit element in $\mathfrak{R}$, if $r = 0$. Stated in words, $A_{\lambda_1 \cdots \lambda_n}$ and $C^{\rho_1 \cdots \rho_q}$ are the determinants $| a_{\lambda_j}^i |$ and $| c_t^{\rho_s} |$ $(i, j = 1, \cdots, n;\ s, t = 1, \cdots, q)$ while $\pm B_{\lambda_1 \cdots \lambda_n}^{\rho_1 \cdots \rho_q}$ is the determinant of the square matrix obtained from $\| b_\rho^\lambda \|$ by striking out the rows $b^{\lambda_1}, \cdots, b^{\lambda_n}$ and the columns $b_{\rho_1}, \cdots, b_{\rho_q}$, unless $\lambda_i = \lambda_j$ or $\rho_s = \rho_t$ for some $i \neq j$ or $s \neq t$, in which case $B_{\lambda_1 \cdots \lambda_n}^{\rho_1 \cdots \rho_q} = 0$. We assume that

$$A_{\lambda_1 \cdots \lambda_n} C^{\rho_1 \cdots \rho_q} = \kappa B_{\lambda_1 \cdots \lambda_n}^{\rho_1 \cdots \rho_q}, \tag{2.6}$$

for some $\kappa \in \mathfrak{R}$, and all values of the indices λ_i, ρ_s. Then it follows from the tensorial character of the components A, B and C that this condition is invariant under transformations of the form

$$\mathbf{a}' = \mathbf{f}\mathbf{a}\mathbf{g}^{-1}, \qquad \mathbf{b}' = \mathbf{g}\mathbf{b}\mathbf{h}^{-1}, \qquad \mathbf{c}' = \mathbf{h}\mathbf{c}\mathbf{k}^{-1},$$

where **f**, **g**, **h** and **k** are regular matrices of appropriate degrees. Moreover, the condition will be satisfied after such a transformation if κ is replaced by $\kappa' = \pi\kappa$, where π is a product of the determinants $f = | f_j^i |$, $g = | g_\mu^\lambda |$, $h = | h_\sigma^\rho |$, $k = | k_t^s |$ and their inverses (actually $\pi = f^{-1} g h^{-1} k$). Clearly the determinant of a regular matrix is a regular element of $\mathfrak{R}$, and the determinant of an elementary $\mathfrak{G}$-matrix is an element of $\mathfrak{G}$. Therefore the multiplier π is a regular element, and $\pi \in \mathfrak{G}$ if **f**, **g**, **h** and **k** are elementary matrices. Finally it is easy to see that the condition (2.6) is invariant under an elementary expansion (not an extended expansion) of the matrices **a**, **b** and **c**, the factor κ being unaltered. Therefore we have the lemma:

LEMMA 5. *If the matrices* **a**, **b**, **c** *satisfy a condition of the form* (2.6), *so do any matrices* **a**′, **b**′, **c**′, *which are L^*-equivalent to* **a**, **b**, **c**, *and with κ replaced by $\kappa' = \pi\kappa$,*

[10] Journal für die Math. r. u. a. (loc. cit.), §6. See also W. Franz, idem., 176 (1937), 113-34.

[11] See, for example, O. Veblen, Invariants of quadratic differential forms, Cambridge (1933), chap. 1.

where π is a regular element of $\mathfrak{A}$. If $\mathbf{a}'$, $\mathbf{b}'$, $\mathbf{c}'$ *are L-equivalent to* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, *then we may take* $\pi \in \mathfrak{G}$.

Clearly κ is uniquely determined by (2.6), unless there is a non-zero $\lambda \in \mathfrak{R}$, such that $\lambda B^{\rho_1 \cdots \rho_q}_{\lambda_1 \cdots \lambda_n} = 0$ for all values of the indices. The condition that there shall, or shall not, exist such a λ is obviously an invariant of L^*-equivalence. Therefore, if κ is uniquely determined, so is κ', and π is uniquely determined unless κ is a 0-divisor, including 0 itself among the 0-divisors. In particular, κ is uniquely determined if $r = 0$.

3. In this and the next section we study the relation between the algebra in §2 and the topology of finite, connected complexes. We first correct the error in Q.J. Let K_1^n and K_2^p $(n \geqq p)$ be two finite, connected simplicial complexes and let $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^p, \cdots, \mathbf{s}^1$ be incidence matrices for K_1^n and K_2^p, which are determined by Reidemeister's process. The elements of $\mathbf{r}^j$ and $\mathbf{s}^k$ are in the group rings, $\mathfrak{R}_1$, and $\mathfrak{R}_2$, of K_1^n and K_2^p. The (integral) group ring $\mathfrak{R}$, of any group $\mathfrak{H}$, satisfies the condition concerning k and l in (2.3). For a homomorphism of $\mathfrak{R}$ on the coefficient ring of $\mathfrak{R}$, in this case the ring of integers, is determined by writing $x = 1$ for each $x \in \mathfrak{H}$. Returning to K_1^n and K_2^p, if $n > p$ we add empty matrices $\mathbf{s}^n, \cdots, \mathbf{s}^{p+1}$ to the set $\mathbf{s}^p, \cdots, \mathbf{s}^1$ and agree that, if $p < q \leqq n$, then a primary expansion of $\mathbf{s}^q$ shall be an elementary expansion, not an extended expansion. Let K_1^n and K_2^p be of the same homotopy type. Then $\pi_1(K_1^n)$ is isomorphic to $\pi_1(K_2^p)$, and by a *special isomorphism* of $\mathfrak{R}_1$ on $\mathfrak{R}_2$ we shall mean one which is determined by an isomorphism of $\pi_1(K_1^n)$ on $\pi_1(K_2^p)$. In the definition of elementary $\mathfrak{G}$-matrices, and L-equivalence of the incidence matrices for K_1^n, we now take $\mathfrak{G}$ to be the group consisting of all the elements $\pm x$, where $x \in \pi_1(K_1^n)$.

Let $\mathfrak{R}_1$ and $\mathfrak{R}_2$ be two isomorphic rings and let $\psi(\mathfrak{R}_1) = \mathfrak{R}_2$ be an isomorphism of $\mathfrak{R}_1$ on $\mathfrak{R}_2$. If $\mathbf{r} = \| r_{i\lambda} \|$, where $r_{i\lambda} \in \mathfrak{R}_1$, and if $r'_{i\lambda} = \psi(r_{i\lambda})$, we shall denote the matrix $\| r'_{i\lambda} \|$ by $\psi(\mathbf{r})$. We shall describe two sets of matrices $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$, whose elements belong to $\mathfrak{R}_1$ and $\mathfrak{R}_2$ respectively, as *X-equivalent under* ψ $(X = L$ or $L^*)$ if, and only if, the matrices $\psi(\mathbf{r}^n), \cdots, \psi(\mathbf{r}^1)$ and $\mathbf{s}^n, \cdots, \mathbf{s}^1$ are X-equivalent. The terms EL-equivalent and L^m-equivalent under ψ will have a similar meaning. What does follow from Q.J., is that the incidence matrices of K_1^n and K_2^p are EL-equivalent under some special isomorphism $\psi(\mathfrak{R}_1) = \mathfrak{R}_2$ if K_1^n and K_2^p have the same nucleus. It then follows from S.S., theorems 17 and 13 (pp. 277, 269) that the incidence matrices of K_1^n and K_2^p are L^n-equivalent if K_1^n and K_2^p are of the same homotopy type. Hence, theorems 1 and 2 lead to the following theorem, which is also a corollary of theorems 6 and 8 in §4 below.

THEOREM 3. *If K_1^n and K_2^p have the same nucleus their incidence matrices are L-equivalent under some special isomorphism $\psi(\mathfrak{R}_1) = \mathfrak{R}_2$. If K_1^n and K_2^p are of the same homotopy type their incidence matrices are L^*-equivalent under a special isomorphism $\psi(\mathfrak{R}_1) = \mathfrak{R}_2$.*

In either case it follows from induction on the number of elementary trans-

formations by which one passes from K_1^n to K_2^p that the isomorphism ψ may be taken from the class which is determined by some homotopy isomorphism[12] $f(K_1^n) \subset K_2^p$.

Theorem 3 states that the geometrical operations described in S.S. can be copied algebraically by the operations described in §2 above. We now turn to the converse question: "to what extent can the algebraic transformations be copied geometrically?" In order to study this question we introduce what we call membrane complexes, which are, so to speak, more elastic than simplicial complexes. Before defining a membrane complex, we modify the definition of a simple membrane, given in S.S., p. 263, by including among the simple membranes, first an n-element E^n, and secondly any stellar subdivision $\sigma\mathbf{E}^n$, as a simple membrane, where $\mathbf{E}^n$ is any simple membrane. If $\mathbf{E}^n = E^n$, an n-element, then its boundary[13] $F(\mathbf{E}^n)$, shall be the identical map of $\dot{E}^n$ on itself. Let $F(\mathbf{E}^n)$ be a simplicial map $f(S^{n-1}) \subset K$, where $\mathbf{E}^n$ is a given membrane, and let (A, a) be an elementary sub-division of $\mathbf{E}^n$. If the simplex A is internal to $\mathbf{E}^n$ then $F(\sigma\mathbf{E}^n) = f(S^{n-1})$, where σ is the subdivision (A, a). If $A \subset F(\mathbf{E}^n)$, let σ_1 be the sub-division of S^{n-1} in which all the simplexes in $f^{-1}(A)$ are starred. Then $F(\sigma\mathbf{E}^n)$ shall be the simplicial map $f'(\sigma_1 S^{n-1})$, in which $f'(b) = a$ if b is one of the new vertices introduced by σ_1, and $f'(b) = f(b)$ if b is one of the original vertices in S^{n-1}. By reiterating this construction we define $F(\sigma\mathbf{E}^n)$, where σ is any stellar sub-division. It is obvious that this extension of the definition does not invalidate any of the results in S.S.

An n-dimensional membrane complex $\mathbf{K}^n$, will be defined by induction on n. A 0-dimensional membrane is a 0-simplex and a 0-dimensional membrane complex is a collection of 0-simplexes. We assume that an $(n-1)$-dimensional complex $\mathbf{K}^{n-1}$, has been defined $(n \geqq 1)$, and that it is a simplicial complex of at most $n-1$ dimensions, whose simplexes are oriented and grouped together to form certain oriented simple membranes, which we shall describe as the *cells* of $\mathbf{K}^{n-1}$. Then an *n-dimensional membrane complex* is to be a complex of the form[14]

$$\mathbf{K}^n = \mathbf{K}^{n-1} + \mathbf{E}_1^n + \cdots + \mathbf{E}_{\alpha_n}^n,$$

where $\mathbf{E}_i^n$ is an oriented, simple membrane, whose boundary is a simplicial, spherical map $F(\mathbf{E}_i^n) \subset \mathbf{K}^{n-1}$. Thus $\mathbf{K}^n$ is a simplicial complex K^n, whose simplexes are grouped together in a particular way to form the cells of $\mathbf{K}^n$.

[12] See note III_1.

[13] As in S.S., $F(\mathbf{E}^p)$ will always denote the "spherical" boundary of a simple membrane $\mathbf{E}^p$. $F(\mathbf{E}^p)$ is a (simplicial) map, not a complex. We shall use $\dot{E}^n$ to denote the boundary of an n-element E^n. The cycle in $\mathbf{K}^{p-1}$, defined below, which is determined by $F(\mathbf{E}_i^p)$, will be denoted by $\partial\mathbf{E}_i^p$. In fact $\partial\mathbf{E}_i^p$ will only be used for membranes in a universal covering complex of a given complex.

[14] As a matter of convention we allow this set of membranes to be empty, in which case $\mathbf{K}^n = \mathbf{K}^{n-1}$. Thus $\mathbf{K}^n$ stands for a complex of at most n dimensions in the ordinary sense of the word, and it is to be assumed throughout this paper that $n > 2$. As in S.S., it is to be assumed that, in the above expression for $\mathbf{K}^n$, or in any similar expression, none of the membranes has in inner point in common with any of the others or with $\mathbf{K}^{n-1}$.

We shall call $'K^n$, or any stellar sub-division of K^n, a *simplicial sub-division* of $\mathbf{K}^n$ and shall say that K^n covers $\mathbf{K}^n$. We shall use $\mathbf{K}^p$ $(0 \leqq p \leqq n)$ to stand for the membrane complex consisting of the cells in $\mathbf{K}^n$ whose dimensionalities do not exceed p, and K^p will stand for the sub-complex of K^n which covers $\mathbf{K}^p$ (not for the p-dimensional skeleton of K^n). Notice that $F(\mathbf{E}_i^p)$, where $\mathbf{E}_i^p \subset \mathbf{K}^p$, though it is a simplicial map in K^{p-1}, need not cover an exact sum of cells in $\mathbf{K}^{p-1}$.

It follows from the definition in S.S. that a simple membrane is a combinatorial cell,[15] and from an argument in S.S., lemma 5, that, if K^{q+1} is a simplicial sub-division of $\mathbf{K}^{q+1}$, then any p-chain $(p \leqq q)$ in K^{q+1} can be deformed into $\mathbf{K}^q$, with the part in $\mathbf{K}^q$ held fixed. Therefore the p-cells $\mathbf{E}_i^p \subset \mathbf{K}^p$ constitute an homology basis[16] for the p-cycles in any simplicial subdivision of $\mathbf{K}^n$. It is to be understood that any 'given complex' is finite and connected, but we note that these definitions apply equally well to infinite complexes. In particular a universal covering complex $\tilde{\mathbf{K}}^n$, of $\mathbf{K}^n$, is obviously a membrane complex, whose cells are the oriented membranes which cover the membranes of $\mathbf{K}^n$. It is always to be understood that the (oriented) cells of $\tilde{\mathbf{K}}^n$ cover the cells of $\mathbf{K}^n$ positively.

We shall restrict the formal deformations of a membrane complex to deformations in which the membranes may be regarded as undivided cells,[17] and the following arguments show that the main results in S.S. are applicable to the resulting scheme. Let $\mathbf{K}$ be a given membrane complex and let $\mathbf{E}^p$ be a new membrane bounded by a simplicial spherical map $F(\mathbf{E}^p) \subset \mathbf{K}$, which is homotopic to a point in $\mathbf{K}$. Let $f(S^p) \subset \mathbf{K} + \mathbf{E}^p$ be a simplicial map which is *simple* in $\mathbf{E}^p$, meaning that there is a p-simplex $A^p \subset E^p$, such that $f^{-1}(A^p)$ is a single simplex. Let $\mathbf{E}^{p+1}$ be a new membrane which is bounded by $f(S^p)$, and let $\mathbf{K}_1 = \mathbf{K} + \mathbf{E}^{p+1}$. Then the transformation $\mathbf{K} \rightarrow \mathbf{K}_1$ will be called an *elementary expansion* of the membrane complex $\mathbf{K}$, and its inverse will be called an *elementary contraction*. By a *formal deformation* D we shall mean a finite sequence of elementary expansions and contractions, whose resultant is a transformation $\mathbf{K} \rightarrow D(\mathbf{K})$. Two membrane complexes $\mathbf{K}_0$ and $\mathbf{K}$ will be said to have the same *nucleus* if, and only if, $\mathbf{K} = D(\mathbf{K}_0)$, where D is some formal deformation. Similarly we take over the definition of an m-group from S.S., defining an elementary filling of order p as a transformation of the form $\mathbf{K} \rightarrow \mathbf{K} + \mathbf{E}^p$, where $\mathbf{E}^p$ is a new membrane whose boundary is an arbitrary simplicial map in $\mathbf{K}^{p-1}$.

Theorem 5 in S.S. is obviously valid for membrane complexes, on the understanding that $\sigma\mathbf{K}$ is the membrane complex which is obtained from $\mathbf{K}$ by a sub-division σK, where K is a simplicial sub-division of $\mathbf{K}$. Theorems 11 and 14 in S.S. (pp. 264 and 272) are valid for membrane complexes. Indeed the proof of theorem 11 can be considerably simplified. For if $F(\mathbf{E}_i^n) \subset \mathbf{K}$ $(i = 0, 1)$ and

[15] See, for example, P. Alexandroff and H. Hopf, Topologie, Berlin (1935), p. 245.

[16] That is to say, any p-cycle in $\mathbf{K}^n$ is homologous to a cycle which consists of (integral) multiples of these cells, and if such a p-cycle bounds in $\mathbf{K}^n$ it bounds a chain composed of the $(p + 1)$-dimensional membranes (cf. Alexandroff and Hopf, loc. cit., pp. 245 et seq.).

[17] Cf. note III_2.

if $F(\mathbf{E}_0^n)$ is homotopic in $\mathbf{K}$ to $F(\mathbf{E}_1^n)$, then there is obviously a spherical map $f(S^n) \subset \mathbf{K} + \mathbf{E}_0^n + \mathbf{E}_1^n$, which is simple both in $\mathbf{E}_0^n$ and in $\mathbf{E}_1^n$. Therefore, assuming first that $\mathbf{E}_i^n$ does not meet $\mathbf{K} + \mathbf{E}_j^n (i, j = 0, 1; j \neq i)$ except in $F(\mathbf{E}_i^n)$, it follows that $\mathbf{K} + \mathbf{E}_i^n$ expands into $\mathbf{K} + \mathbf{E}^{n+1}$, where $F(\mathbf{E}^{n+1}) = f(S^n)$. Therefore

$$\mathbf{K} + \mathbf{E}_1^n = D(\mathbf{K} + \mathbf{E}_0^n) \qquad (\text{rel. } \mathbf{K}),$$

where 'rel. **K**' means the same as in S.S. If $\mathbf{E}_0^n$ and $\mathbf{E}_1^n$ have internal intersections, we still have

$$\mathbf{K} + \mathbf{E}_i^n = D(\mathbf{K} + \mathbf{E}_2^n) \qquad (\text{rel. } \mathbf{K}),$$

where $F(\mathbf{E}_2^n) = F(\mathbf{E}_0^n)$ and $\mathbf{E}_2^n$ does not meet $\mathbf{K} + \mathbf{E}_0^n + \mathbf{E}_1^n$ except in $F(\mathbf{E}_2^n)$. This establishes the generalization of theorem 11, and the generalization of theorem 14 follows from the arguments given in S.S. We shall refer to these generalizations simply as S.S., theorems 11 and 14.

Let $\mathbf{K} \to \mathbf{K}_1 = \mathbf{K} + \mathbf{E}^{n+1}$ be an elementary expansion of $\mathbf{K}$, where the map $F(\mathbf{E}^{n+1}) \subset \mathbf{K} + \mathbf{E}^n$ is simple in the membrane $\mathbf{E}^n$, which does not meet $\mathbf{K}$ except in $F(\mathbf{E}^n) \subset \mathbf{K}$. Then I say that K expands geometrically into K_1, in the sense of S.S., §5 (p. 258), where K_1 is a simplicial sub-division of $\mathbf{K}_1$ and K is the sub-complex of K_1 which covers $\mathbf{K}$. For, first let $\mathbf{E}^n = E^n$, an n-element, and let $A^n \subset E^n$ be an n-simplex, which has but a single original in the map $F(\mathbf{E}^{n+1})$. After a suitable sub-division of E^n, we may assume that all the vertices of A^n are internal to E^n, and the assertion follows from S.S., theorem 1 (p. 250), the corollary to theorem 8 (p. 260) and lemma 6 (p. 265). If $\mathbf{E}^n$ is of the form $E^n + C_f(\dot{E}^n)$, as defined in S.S., p. 263, then $\mathbf{E}^n - \mathbf{A}^n$ contracts into $F(\mathbf{E}^n)$, where $\mathbf{A}^n$ is any open simplex in $\mathbf{E}^n$. For $\mathbf{E}^n = E_0^n + N$, where E_0^n is an n-element which contains[18] E^n in its interior, and N is the set of closed simplexes in $\mathbf{E}^n$, which meet $F(\mathbf{E}^n)$. If $\mathbf{A}^n \in E^n$ it follows that $\mathbf{E}^n - \mathbf{A}^n$ contracts geometrically into N, and, from an argument similar to the proof of theorem 8 in S.S., that N contracts into $F(\mathbf{E}^n)$. If $\mathbf{A}^n \in C_f(\dot{E}^n)$ it follows from an argument similar to one on p. 281 of S.S. that $\mathbf{E}^n - \mathbf{A}^n$ contracts into $F(\mathbf{E}^n)$. Finally, if $\mathbf{E}^n = \sigma\{E^n + C_f(\dot{E}^n)\}$ it follows from what we have just proved and S.S., lemma 7 (p. 267), that $\mathbf{E}^n - \mathbf{A}^n$ contracts geometrically into $F(\mathbf{E}^n)$, where $\mathbf{A}^n$ is any open simplex in $\mathbf{E}^n$. Therefore it follows from S.S., lemma 6, that $\mathbf{K}$ expands geometrically into $\mathbf{K}$, and the assertion is justified in each case. Hence, and from S.S., lemma 5 (p. 265), it follows that, if $\mathbf{K}_0$ and $\mathbf{K}_1$ have the same nucleus or m-group, for a given value of m, then so do any simplicial sub-divisions of $\mathbf{K}_0$ and $\mathbf{K}_1$. The converse, and also the generalization of theorem 17 in S.S. p. 277, follows from the theorem:

THEOREM 4. *If K is a simplicial sub-division of* $\mathbf{K}$, *then* $\mathbf{K} = D(K)$.

If $\mathbf{K} = D_0(K^*)$, where K^* is any simplicial complex (i.e. a membrane complex whose cells are simplexes), we have $K = D_1(K^*)$, where K is a simplicial sub-division of $\mathbf{K}$, whence $\mathbf{K} = D_0 D_1^{-1}(K)$. Therefore the theorem will follow when

[18] E_0^n may be regarded as a triangulation of $E^n + (\dot{E}^n \times \langle 0, \tfrac{1}{2} \rangle)$ (Cf. S.S., p. 259).

we have proved that $\mathbf{K} = D_0(K^*)$, where K^* is a simplicial complex. This is trivial if $\mathbf{K}$ consists of a single vertex, and we shall prove it by induction on the number of cells in $\mathbf{K}$. Let $\mathbf{K} = \mathbf{K}_0 + \mathbf{E}_1^n$, where $\mathbf{E}_1^n$ is a cell of maximum dimensionality, and $\mathbf{K}_0$ consists of the remaining cells in $\mathbf{K}$. By the hypothesis of the induction, there is a simplicial complex K_0, such that $K_0 = D_0(\mathbf{K}_0)$, and by S.S., theorem 14, D_0 may be extended to a deformation $\mathbf{K} \to D_1(\mathbf{K}) = K_0 + \mathbf{E}^n = \mathbf{K}_1$, say, where $\mathbf{E}^n = D_1(\mathbf{E}_1^n)$, and $\mathbf{K}_1$ consists of the simplexes in K_0, together with the cell $\mathbf{E}^n$. Let $\mathbf{E}_0^n$ be a new membrane with the same boundary as $\mathbf{E}^n$, which does not meet $\mathbf{K}_1$ except in $F(\mathbf{E}_0^n) \subset K_0^{n-1}$, and let $\mathbf{A}^n \epsilon \mathbf{E}_0^n$ be an open n-simplex, whose closure is internal to $\mathbf{E}_0^n$. Let $K^* = K_0 + \mathbf{E}_0^n$ be the simplicial complex, whose cells are the simplexes in K_0 and in $\mathbf{E}_0^n$, and let $\mathbf{E}^{n+1}$ be a new membrane, whose boundary is the n-sphere[19] $\mathbf{E}^n - \mathbf{E}_0^n$. Then $F(\mathbf{E}^{n+1})$ is simple in $\mathbf{A}^n$, whence $\mathbf{K}_1$ expands first into $\mathbf{K}_1 + \mathbf{E}_0^n - \mathbf{A}^n$, and then into $\mathbf{K}_1 + \mathbf{E}^{n+1}$. But $F(\mathbf{E}^{n+1})$ is also simple in $\mathbf{E}^n$, whence $\mathbf{K}_1 + \mathbf{E}^{n+1}$ contracts into K^*, and the theorem is established.

We now proceed to the algebraic description of a (membrane) complex $\mathbf{K}^n$. Let $\tilde{\mathbf{K}}^n$ be a universal covering complex of $\mathbf{K}^n$, and $u(\tilde{\mathbf{K}}^n) = \mathbf{K}^n$ a locally $(1-1)$ map of $\tilde{\mathbf{K}}^n$ on $\mathbf{K}^n$. Let $\mathbf{E}^0 = \mathbf{E}_1^0$, say, be an arbitrary vertex of $\mathbf{K}^n$ and let $\tilde{\mathbf{E}}^0 \epsilon \tilde{\mathbf{K}}^n$ be an arbitrary vertex in $u^{-1}(\mathbf{E}^0)$. We take $\mathbf{E}^0$ as the base point for $\pi_1(\mathbf{K}^n)$, and identify each element in the covering group[20] of $\tilde{\mathbf{K}}^n$ with the element of $\pi_1(\mathbf{K}^n)$ to which it corresponds in the isomorphism determined by $\mathbf{E}^0$ and $\tilde{\mathbf{E}}^0$. Then an arbitrary vertex in $u^{-1}(\mathbf{E}^0)$ will be denoted by $x\tilde{\mathbf{E}}^0$, where $x \epsilon \pi_1(\mathbf{K}^n)$. Let $\mathbf{V}$ be a (connected) tree which contains $\mathbf{K}^0$ and is a sub-complex of $\mathbf{K}^1$, and let $x\tilde{\mathbf{V}}$ be the component of $u^{-1}(\mathbf{V}) \subset \tilde{\mathbf{K}}^n$ which contains $x\tilde{\mathbf{E}}^0$. Let $\mathbf{E}_1^1, \cdots, \mathbf{E}_k^1$ be the oriented 1-cells of $\mathbf{K}^1$, if there are any, which are not in $\mathbf{V}$, and let $x\tilde{\mathbf{E}}_i^1$ be the 1-cell of $\tilde{\mathbf{K}}^n$ which covers $\mathbf{E}_i^1$ positively and whose initial point is on $x\tilde{\mathbf{V}}$. In calculating the incidence matrices $\mathbf{r}^2$, $\mathbf{r}^1$ for $\mathbf{K}^n$, we shall treat $\mathbf{V}$ as if it were a single point, ignoring the 1-cells in $\mathbf{V}$ and the vertices other than $\mathbf{E}^0$. On this understanding, we shall take[21] $\tilde{\mathbf{V}} = 1\tilde{\mathbf{V}}$ and $\tilde{\mathbf{E}}_i^1 = 1\tilde{\mathbf{E}}_i^1$ as basis elements for the 0-chains and 1-chains in $\tilde{\mathbf{K}}^n$. Similarly we shall treat $\mathbf{V}$ as if it were a single point in setting up a system of generators and relations for $\pi_1(\mathbf{K}^n)$. Thus any segment $s \subset \mathbf{K}^1$, which begins and ends on $\mathbf{V}$, is to be automatically transformed into the circuit $s_0 + s - s_1$, where s_0 and s_1 are segments on $\mathbf{V}$ which join $\mathbf{E}^0$ to the first and last points of s, and $s_i (i = 0, 1)$ is non-singular unless it degenerates into $\mathbf{E}^0$. Such a circuit is determined, up to homotopic deformation over $\mathbf{K}^1$, by an expression of the form

$$s = \epsilon_1 \mathbf{E}_{i_1}^1 + \cdots + \epsilon_q \mathbf{E}_{i_q}^1 \qquad (\epsilon_\lambda = \pm 1), \tag{3.1}$$

in which the summation is non-commutative and the last vertex of $\epsilon_\lambda \mathbf{E}_{i_\lambda}^1$ is joined to the first vertex of $\epsilon_{\lambda+1}\mathbf{E}_{i_{\lambda+1}}^1$ by a segment on $\mathbf{V}$. The corresponding segment

[19] Cf. S.S., p. 264.

[20] I.e. the group of homeomorphisms (covering transformations) $t(\tilde{\mathbf{K}}^n) = \tilde{\mathbf{K}}^n$, such that $ut = u$.

[21] We now denote the unit element in any group, and in its group ring, by 1.

in $\tilde{\mathbf{K}}^n$, which starts at $\tilde{\mathbf{V}}$, is given by an expression of the form

$$\tilde{s} = \epsilon_1 x_1 \tilde{\mathbf{E}}^1_{i_1} + \cdots + \epsilon_q x_q \tilde{\mathbf{E}}^1_{i_q}, \tag{3.2}$$

where $x_\lambda \in \pi_1(\mathbf{K}^n)$. The circuit (3.1) and the segment (3.2) are also represented by the product

$$x = a_{i_1}^{\epsilon_1} \cdots a_{i_q}^{\epsilon_q}, \tag{3.3}$$

where a_i is the (free) generator of the free group $\mathfrak{F}_k = \pi_1(\mathbf{K}^1)$, which corresponds to the 'circuit' $\mathbf{E}^1_i$. Let $\mathfrak{R}$ be the group ring of $\mathbf{K}^n$, and let

$$C^1 = r_1 \tilde{\mathbf{E}}^1_1 + \cdots + r_k \tilde{\mathbf{E}}^1_k \qquad (r_i \in \mathfrak{R}) \tag{3.4}$$

be the chain which is obtained by simplifying (3.2), with commutative addition. Since $\tilde{s}$ is an unbroken segment we have

$$\partial C^1 = (x - 1)\tilde{\mathbf{V}}, \tag{3.5}$$

where ∂C denotes the boundary of a given chain C. In particular $\partial \tilde{\mathbf{E}}^1_i = (a_i - 1)\tilde{\mathbf{V}}$, whence $\mathbf{r}^1$ is the matrix with a single column and $a_i - 1$ in the i^{th} row, if $k > 0$. If $k = 0$, then $\mathbf{r}^1$ has one column and no rows.

We now calculate the coefficients x_λ in (3.2) in terms of the product (3.3). Let $\tilde{s}_\lambda$ be the part of $\tilde{s}$ which is given by $y_\lambda = a_{i_1}^{\epsilon_1} \cdots a_{i_\lambda}^{\epsilon_\lambda}$, with $\tilde{s}_0 = \tilde{\mathbf{E}}^0$, $y_0 = 1$. Then the last vertex of $\tilde{s}_{\lambda-1}$, and hence the first vertex of $\epsilon_\lambda x_\lambda \tilde{\mathbf{E}}^1_i$, are on $y_{\lambda-1}\tilde{\mathbf{V}}$. Therefore, if $\epsilon_\lambda = +1$, the first vertex of $x_\lambda \tilde{\mathbf{E}}^1_{i_\lambda}$ is on $y_{\lambda-1}\tilde{\mathbf{V}}$, whence $x_\lambda = y_{\lambda-1}$. If $\epsilon_\lambda = -1$ the first vertex of $x_\lambda \tilde{\mathbf{E}}^1_{i_\lambda}$ is on $y_{\lambda-1}a_{i_\lambda}^{-1}\tilde{\mathbf{V}}$, in which case $x_\lambda = y_{\lambda-1}a_{i_\lambda}^{-1}$. Therefore[22]

$$\begin{aligned} x_\lambda &= a_{i_1}^{\epsilon_1} \cdots a_{i_{\lambda-1}}^{\epsilon_{\lambda-1}} && \text{if } \epsilon_\lambda = +1 \\ &= a_{i_1}^{\epsilon_1} \cdots a_{i_\lambda}^{\epsilon_\lambda} && \text{if } \epsilon_\lambda = -1. \end{aligned} \tag{3.6}$$

Notice that, in either case,

$$\begin{aligned} \epsilon_\lambda x_\lambda \partial \tilde{\mathbf{E}}^1_{i_\lambda} &= y_{\lambda-1}(a_{i_\lambda}^{\epsilon_\lambda} - 1)\tilde{\mathbf{V}} \\ &= (y_\lambda - y_{\lambda-1})\tilde{\mathbf{V}}. \end{aligned}$$

Therefore, if ∂ is an abstract operator such that $\partial \tilde{\mathbf{E}}^1_i = (a_i - 1)\tilde{\mathbf{V}}$ and $\partial(r_1 C^1_1 + r_2 C^1_2) = r_1 \partial C^1_1 + r_2 \partial C^1_2$, for any pair of chains C^1_1, C^1_2, and elements $r_1, r_2 \subset \mathfrak{R}$, then (3.5) follows by a purely formal calculation from (3.2) and

[22] In formal calculations which involve (3.6) one must remember the ambiguity in the notation, which is referred to in note III_5. Thus (3.3), as an alternative expression for (3.1) or (3.2), is simply a product of the generators. On the other hand the coefficients in (3.2) are elements of $\pi_1(\mathbf{K}^n)$, and x_λ may be replaced by any product of the generators which represents the same element. For example, it may be verified formally that $C^1 = 0$ if $x = 1$ in $\mathfrak{F}_k$, where x and C^1 are given by (3.3) and (3.4), and hence that, if $x_2 = yx_1y^{-1}$, where $x_1 = 1$ in $\pi_1(\mathbf{K}^n)$, and if C^1_i is the chain which corresponds to x_i ($i = 1, 2$), then $C^1_2 = yC^1_1$ (cf. (4.2), below).

(3.3). Notice also that

(3.7) $$r_1(a_1 - 1) + \cdots + r_k(a_k - 1) = x - 1$$

in consequence of (3.4) and (3.5).

Let $\mathbf{E}_1^2, \cdots, \mathbf{E}_{\alpha_2}^2$ be the 2-cells of $\mathbf{K}^n$, if there are any, and let $f_\lambda(S_\lambda^1) \subset \mathbf{K}^1$ be a map which is equivalent to the spherical boundary $F(\mathbf{E}_\lambda^2)$. If $f_\lambda(S_\lambda^1) \subset \mathbf{K}^1 - \mathbf{K}^0$ for some value of λ, then $f_\lambda(S_\lambda^1)$ is internal to some 1-cell in $\mathbf{K}^1$, since it is connected, and is therefore homotopic, in $\mathbf{K}^1$, to a point. If not, we choose a base point $p_\lambda \in S_\lambda^1$, such that $f_\lambda(p_\lambda) = \mathbf{E}_\lambda^0 \in \mathbf{V}$, where $\mathbf{E}_\lambda^0$ is any vertex of $\mathbf{K}^0$. In either case $F(\mathbf{E}_\lambda^2)$, with the base point $f_\lambda(p_\lambda)$ if $F(\mathbf{E}_\lambda^2) \not\subset \mathbf{K}^1 - \mathbf{K}^0$, determines an element $R_\lambda \in \mathfrak{F}_k$, where R_λ is the element $1 \in \mathbf{F}_k$ if $F(\mathbf{E}_\lambda^2) \subset \mathbf{K}^1 - \mathbf{K}^0$. Then $\pi_1(\mathbf{K}^n)$ is generated by $a_1, \cdots, a_k$, subject to the relations $R_1 = \cdots = R_{\alpha_2} = 1$. We shall denote this system of generators and relations by R, and shall call it a *natural system*,[23] of generators and relations for $\pi_1(\mathbf{K}^n)$. It is uniquely determined by the choice of V, the choice of orientations for $\mathbf{E}_i^1$ and $\mathbf{E}_\lambda^2$ and of the base points $p_\lambda \in S_\lambda^1$. If $F(\mathbf{E}_\lambda^2) \not\subset \mathbf{K}^1 - \mathbf{K}^0$ let A_λ^2 be a closed simplex in $\mathbf{E}_\lambda^2$, which meets $F(\mathbf{E}_\lambda^2)$ in the base point $f_\lambda(p_\lambda)$ and nowhere else. Let $\tilde{A}_\lambda^2 \subset \tilde{\mathbf{K}}^2$ be the simplex in $u^{-1}(A_\lambda^2)$ which meets $\tilde{\mathbf{V}}$, and let $\tilde{\mathbf{E}}_\lambda^2 \subset \tilde{\mathbf{K}}^2$ be the cell in $u^{-1}(\mathbf{E}_\lambda^2)$ which contains $\tilde{A}_\lambda^2$, or any cell in $u^{-1}(\mathbf{E}_\lambda^2)$ if $F(\mathbf{E}_\lambda^2) \subset \mathbf{K}^1 - \mathbf{K}^0$. Then the chain boundary $\partial\tilde{\mathbf{E}}_\lambda^2$ is given by (3.4) when R_λ is given by (3.3). Therefore the incidence matrix $\mathbf{r}_\lambda^2$ can be calculated algebraically from the system R.

Let $\tilde{\mathbf{E}}_i^q \subset \tilde{\mathbf{K}}^q$ be an arbitrary cell in $u^{-1}(\mathbf{E}_i^q)$ for each $q = 3, \cdots, n$ and $i = 1, \cdots, \alpha_q$. Since the cells $\tilde{\mathbf{E}}_\lambda^{p-1}$ $(\lambda = 1, \cdots, \alpha_{p-1}; p > 2)$ constitute an homology basis, with coefficients in $\mathfrak{R}$, for the $(p-1)$-cycles in $\tilde{\mathbf{K}}^{p-1}$, and since homology implies equality in the top dimension, it follows that the chain boundaries $\partial\tilde{\mathbf{E}}_i^p$ are given by equations of the form

$$\partial\tilde{\mathbf{E}}_i^p = \sum_{\lambda=1}^{\alpha_{p-1}} r_{i\lambda}^p \tilde{\mathbf{E}}_\lambda^{p-1} \qquad (r_{i\lambda}^p \in \mathfrak{R}).$$

We shall describe the matrices $\mathbf{r}^p = \| r_{i\lambda}^p \|$ $(p = 3, \cdots, n)$, together with the matrices $\mathbf{r}^2$, $\mathbf{r}^1$ which are determined by R, as *natural incidence matrices* for $\mathbf{K}^n$. Since $\mathbf{r}^2$ and $\mathbf{r}^1$ are determined by R, the system $(\mathbf{r}, R)$, which consists of the incidence matrices $\mathbf{r}^n, \cdots, \mathbf{r}^3$, together with R, provides at least as precise, and possibly[24] a more precise description of $\mathbf{K}^n$ than that given by the natural incidence matrices. We shall call $(\mathbf{r}, R)$ a *natural system* for $\mathbf{K}^n$.

We now investigate the conditions under which two complexes may be described by the same natural system. The following arguments lead up to the statement that a natural system for a given complex $\mathbf{K}_0^n$, is unaltered by a

[23] Any system of generators and relations is a natural system for some complex $\mathbf{K}^2$ (cf. O. Veblen, Analysis Situs, New York (1931), chap. V, §24). Thus the system R may be arbitrary except that, for the purposes of this paper, it shall be finite. In particular there may be no generators and $\alpha_2 \geqq 0$ vacuous relations. Or we may have $\alpha_2 = 0$, $k > 0$.

[24] The question implied is: "to what extent is $\mathbf{K}^2$ determined by an abstract knowledge of $\pi_1(\mathbf{K}^2)$, together with given expressions for $\mathbf{r}^2$, $\mathbf{r}^1$, in terms of an arbitrary representation of $\pi_1(\mathbf{K}^2)$?"

'trivial deformation' defined below of $\mathbf{K}_0^n$. Let $\mathbf{K}_0^n$ be a given complex ($n > 0$) and let

$$\mathbf{K}_0^p \to \mathbf{K}_1^p = D_p(\mathbf{K}_0^p) \qquad (0 < p < n)$$

be any formal deformation of $\mathbf{K}_0^p$ into another p-dimensional complex $\mathbf{K}_1^p$. It follows from induction on q ($p < q \leqq n$) and S.S., theorem 14, that the deformation D_p can be extended to a deformation

$$\begin{aligned}\mathbf{K}_0^q \to \mathbf{K}_1^q &= D_q(\mathbf{K}_0^q)\\ &= \mathbf{K}_1^{q-1} + \mathbf{E}_{11}^q + \cdots + \mathbf{E}_{1\alpha_q}^q,\end{aligned}$$

where

$$\mathbf{K}_0^q = \mathbf{K}_0^{q-1} + \mathbf{E}_{01}^q + \cdots + \mathbf{E}_{0\alpha_q}^q$$

and $\mathbf{E}_{1\lambda}^q = D_q(\mathbf{E}_{0\lambda}^q)$. Thus $D_p(\mathbf{K}_0^p)$ can finally be extended to a deformation $D_n(\mathbf{K}_0^n)$, which is also an extension of $D_q(\mathbf{K}_0^q)$ for each $q = p + 1, \cdots, n - 1$. By S.S., theorem 5, there is a complex $\mathbf{K}_p^*$, which contracts both into $\mathbf{K}_0^p$ and into some subdivision of $\mathbf{K}_1^p$, which, for the purpose of this argument, we may take to be $\mathbf{K}_1^p$ itself. On comparing the proof of S.S., theorem 11, which is given above, with the proof of theorem 14 in S.S., we see that there are complexes $\mathbf{K}_{p+1}^*, \cdots, \mathbf{K}_n^*$, such that

$$\mathbf{K}_q^* = \mathbf{K}_{q-1}^* + C_1^{q+1} + \cdots + C_{\alpha_q}^{q+1} \quad (q = p + 1, \cdots, n),$$

where C_λ^{q+1} is a 'deformation cell' for the formal deformation $\mathbf{E}_{0\lambda}^q \to \mathbf{E}_{1\lambda}^q = D_q(\mathbf{E}_{0\lambda}^q)$. The spherical boundary $F(C_\lambda^{q+1})$ consists of $\mathbf{E}_{0\lambda}^q$, $\mathbf{E}_{1\lambda}^q$ and a deformation cylinder for the homotopic deformation $F(\mathbf{E}_{0\lambda}^q) \to F(\mathbf{E}_{1\lambda}^q)$ in $\mathbf{K}_{q-1}^*$. Then $\mathbf{K}_q^*$ contracts into

$$\mathbf{K}_{q-1}^* + \mathbf{K}_i^q = \mathbf{K}_{q-1}^* + \mathbf{E}_{i1}^q + \cdots + \mathbf{E}_{i\alpha_r}^q \qquad (i = 0, 1),$$

and so, by induction on q, into $\mathbf{K}_i^q$. Therefore there is a projection $f_{q-1}(\mathbf{K}_{q-1}^*) = \mathbf{K}_1^{q-1}$ and a projection $g(\mathbf{K}_q^*) = \mathbf{K}_{q-1}^* + \mathbf{K}_1^q$, for a given value of $q > p$, where a projection means a transformation f, such that $f^2 = f$. Let f_{q-1} be extended throughout $\mathbf{K}_{q-1}^* + \mathbf{K}_1^q$ by writing $f_{q-1}(p) = p$ for each point $p \in \mathbf{K}_1^q$. Then $f_q = f_{q-1}g$ is a projection of $\mathbf{K}_q^*$ on $\mathbf{K}_1^q$, such that $f_q(\mathbf{K}_{q-1}^*) = f_{q-1}(\mathbf{K}_{q-1}^*) = \mathbf{K}_1^{q-1}$, and it follows from induction on q that there is a projection $f(\mathbf{K}_n^*) = \mathbf{K}_1^n$, such that $f(\mathbf{K}_q^*) = \mathbf{K}_1^q$ for each $q = p, \cdots, n$. After a suitable deformation of the map f_p we assume that $f(\mathbf{E}_0^0)$ is a vertex $\mathbf{E}_1^0 \in \mathbf{K}_1^0$, where $\mathbf{E}_0^0$ is a given vertex of $\mathbf{K}_0^0$. Let $\tilde{\mathbf{K}}_n^*$ be the universal covering complex of $\tilde{\mathbf{K}}_n^*$, in which a point $\tilde{p} \in \tilde{\mathbf{K}}_n^*$ is defined as a class of curves joining $\mathbf{E}_1^0$ to a given point $p = u(\tilde{p}) \in \mathbf{K}_n^*$. Since $\mathbf{K}_n^*$ contracts into $\mathbf{K}_i^n$ ($i = 0, 1$), whence the latter is a retract by deformation of $\mathbf{K}_n^*$, it follows that $\tilde{\mathbf{K}}_i^n = u^{-1}(\mathbf{K}_i^n)$ is a universal covering complex of $\mathbf{K}_i^n$. Let $\tilde{f}(\tilde{\mathbf{K}}_n^*) = \tilde{\mathbf{K}}_1^n$ be the projection of $\tilde{\mathbf{K}}_n^*$ on $\tilde{\mathbf{K}}_1^n$ in which a given point $\tilde{p} \in \tilde{\mathbf{K}}_n^*$, represented by a segment s, joining $\mathbf{E}_1^0$ to $u(\tilde{p}) \in \mathbf{K}_n^*$, is transformed into the point which is represented by the segment $f(s)$. If $q = 0$, then $\mathbf{K}_q^*$ consists of isolated simply connected complexes, each of which

contracts into a vertex of $\mathbf{K}_i^0$ $(i = 0, 1)$. Similar remarks apply to $\tilde{\mathbf{K}}_0^* = u^{-1}(\tilde{\mathbf{K}}_0^*)$, and it is clear that $\tilde{f}(\tilde{\mathbf{K}}_0^*) = \tilde{\mathbf{K}}_1^0$. In any case the cells in $\tilde{\mathbf{K}}_n^*$, but not in $\mathbf{K}_q^*$, are at least $(q + 1)$-dimensional, and hence at least 2-dimensional if $q > 0$. Therefore, if $q > 0$, we may take $s \subset \mathbf{K}_q^*$, whence $f(s) \subset \mathbf{K}_1^q$ and again $\tilde{f}(\tilde{\mathbf{K}}_q^*) = \tilde{\mathbf{K}}_1^q$, where $\tilde{\mathbf{K}}_q^* = u^{-1}(\mathbf{K}_q^*)$. Therefore, $\tilde{f}(\tilde{\mathbf{K}}_q^*) = \tilde{\mathbf{K}}_1^q$ for each $q = p, \cdots, n$ and $p \geqq 0$. Since $\mathbf{K}_n^*$ contracts into $\mathbf{K}_1^n$ it follows that any circuit in $\mathbf{K}_n^*$, which begins and ends at $\mathbf{E}_1^0$, can be deformed into $\mathbf{K}_1^n$, with $\mathbf{E}_1^0$ held fixed.[25] Therefore $\tilde{f}t = t\tilde{f}$, where t is any covering transformation of $\tilde{\mathbf{K}}_n^*$.

Since $\tilde{f}(\tilde{\mathbf{K}}_r^*) = \tilde{\mathbf{K}}_1^r$ $(r = q - 1, q; q > p)$, and since homology implies equality between relative q-cycles in $\tilde{\mathbf{K}}_1^q$, mod. $\tilde{\mathbf{K}}_1^{q-1}$, the projection $\tilde{f}$ determines a homomorphism ψ_q, of the group of q-cycles in $\tilde{\mathbf{K}}_q^*$, mod. $\tilde{\mathbf{K}}_{q-1}^*$, on the group of q-cycles in $\tilde{\mathbf{K}}_1^q$, mod. $\tilde{\mathbf{K}}_1^{q-1}$, and $\psi(C^q) = C^q$ if $C^q \subset \tilde{\mathbf{K}}_1^q$. A relative q-cycle in $\tilde{\mathbf{K}}_i^q$, mod. $\tilde{\mathbf{K}}_i^{q-1}$, is simply a q-chain which is composed of the cells in $u^{-1}(\mathbf{E}_{i\lambda}^q)$. Therefore the transformation given by $C^q \rightarrow \psi_q(C^q)$, where C^q is an arbitrary q-chain in $\tilde{\mathbf{K}}_0^q$, is a homomorphism of the group of q-chains in $\tilde{\mathbf{K}}_0^q$ in the group of q-chains in $\tilde{\mathbf{K}}_1^q$.

Let R and R' be natural systems of generators and relations for $\pi_1(\mathbf{K}_0^n)$ and $\pi_1(\mathbf{K}_1^n)$, with $\mathbf{E}_0^0$ and $\mathbf{E}_1^0 = f(\mathbf{E}_0^0)$ taken as base points. Let $\pi_1(\mathbf{K}_i^n)$ $(i = 0, 1)$ be identified with the covering group of $\tilde{\mathbf{K}}_i^n$, by taking $\tilde{\mathbf{E}}_i^0$ as a base point in $\tilde{\mathbf{K}}_i^0$, where $\tilde{\mathbf{E}}_i^0$ corresponds to a segment s_i, joining $\mathbf{E}_1^0$ to $\mathbf{E}_i^0$, such that $f(s_i) \subset \mathbf{K}_1^n$ is homotopic to a point.[26] Then the map $f(\mathbf{K}_0^n) = \mathbf{K}_1^n$ determines a natural isomorphism $\psi\{\pi_1(\mathbf{K}_0^n)\} = \pi_1(\mathbf{K}_1^n)$, in which corresponding elements, regarded as covering transformations of $\tilde{\mathbf{K}}_0^n$ and $\tilde{\mathbf{K}}_1^n$, are the transformations induced by the same covering transformation of $\tilde{\mathbf{K}}_n^*$. Let $x \rightarrow \bar{x} = \phi(x)$ be the isomorphism of the symbolic group[27] $\mathfrak{G}_R$ on the symbolic group $\mathfrak{G}_{R'}$ $(x \in \mathfrak{G}_R, \bar{x} \in \mathfrak{G}_{R'})$, by means of which ψ is expressed algebraically, and let $r \rightarrow \bar{r} = \phi(r)$ be the corresponding isomorphism of $\mathfrak{R}$ on $\mathfrak{R}'$, where $\mathfrak{R}$ and $\mathfrak{R}'$ are the group rings of $\mathfrak{G}_R$ and $\mathfrak{G}_{R'}$. Since $\tilde{f}t = t\tilde{f}$, it follows that $\psi_q t = t\psi_q$, whence

$$\psi_q\{\sum_\lambda r_\lambda \tilde{\mathbf{E}}_{0\lambda}^q\} = \sum_\lambda \bar{r}_\lambda \psi_q(\tilde{\mathbf{E}}_{0\lambda}^q), \qquad (q = p + 1, \cdots, n), \tag{3.8}$$

where the coefficients r_λ and $\bar{r}_\lambda = \phi(r_\lambda)$ are expressed algebraically as elements of $\mathfrak{R}$ and $\mathfrak{R}'$.

Let $\tilde{\mathbf{E}}_{0\lambda}^q$ be the cell in $u^{-1}(\mathbf{E}_{0\lambda}^q)$, which is chosen as a basis element in defining a set of natural incidence matrices for $\mathbf{K}_0^n$. If $q > p$ we have

$$[illegible]_{\lambda}^q - \tilde{\mathbf{E}}_{1\lambda}^q = \partial \tilde{C}_\lambda^{q+1} \qquad (\text{mod. } \tilde{\mathbf{K}}_{q-1}^*),$$

[25] See also S.S., theorem 15. This fact has been referred to, by implication, in the statement that $\tilde{\mathbf{K}}_1^n$ is a (single) universal covering complex of $\mathbf{K}_1^n$.

[26] We may take s_0 to be of the form $-f(s) + s$, where s is any segment which joins $\mathbf{E}_1^0$ to $\mathbf{E}_0^0$.

[27] Cf. note III$_3$. $\mathfrak{G}$ stands for the group defined by a given set of generators and relations R.

where $\tilde{C}_\lambda^{q+1}$ and $\tilde{\mathbf{E}}_{1\lambda}^q$ are certain cells in $u^{-1}(C_\lambda^{q+1})$ and $u^{-1}(\mathbf{E}_{1\lambda}^q)$. Since homology implies equality in the top dimensions, and since $\psi_q(C^q) = C^q$ if $C^q \subset \tilde{\mathbf{K}}_1^q$, we have the equality,

$$\psi_q(\tilde{\mathbf{E}}_{0\lambda}^q) = \tilde{\mathbf{E}}_{1\lambda}^q \qquad (\text{mod. } \tilde{\mathbf{K}}_1^{q-1}),$$

between relative q-cycles, which may be interpreted as the absolute equality

$$(3.9) \qquad \psi_q(\tilde{\mathbf{E}}_{0\lambda}^q) = \tilde{\mathbf{E}}_{1\lambda}^q \qquad (q = p+1, \cdots, n),$$

between chains. If $q > p + 1$ it follows from the relation $\partial\tilde{f} = \tilde{f}\partial$, or $\partial\psi_q = \psi_{q-1}\partial$ and from (3.9) and (3.8), that

$$(3.10) \qquad \begin{aligned} \partial\tilde{\mathbf{E}}_{1\lambda}^q &= \psi_{q-1}(\partial\tilde{\mathbf{E}}_{0\lambda}^q) \\ &= \psi_{q-1}\left(\sum_{a=1}^{\alpha} r_{\lambda a}^q \tilde{\mathbf{E}}_{0a}^{q-1}\right) \qquad (\alpha = \alpha_{q-1}) \\ &= \sum_{a=1}^{\alpha} \bar{r}_{\lambda a}^q \tilde{\mathbf{E}}_{1a}^{q-1} \qquad (q = p+2, \cdots, n). \end{aligned}$$

We express by saying that, if $q > p + 1$, then the deformation D_n transforms the incidence matrix $\mathbf{r}^q$, for $\mathbf{K}_0^n$ into the incidence matrix $\phi(\mathbf{r}^q) = \| \bar{r}_{\lambda a}^q \|$, for $\mathbf{K}_1^n$.

We now state three conditions under which $\mathfrak{G}_R$ is to be identified with $\mathfrak{G}_{R'}$, and in such a way that ϕ is the identical automorphism. Under these conditions we shall say that $\mathfrak{G}_R$ is unaltered by a deformation $D(\mathbf{K}_0^n) = \mathbf{K}_1^n$. First let $D(\mathbf{K}_0^n)$ be relative to $\mathbf{K}_0^1$, as the term is defined in S.S., p. 255. Then the generators in the system R may be taken as the generators in R', and the two sets of relations will be equivalent (i.e. each set will imply the other). Therefore $\mathfrak{G}_R = \mathfrak{G}_{R'}$. Also both $\mathbf{K}_0^1 = \mathbf{K}_1^1$, and $\mathbf{E}_0^0 = \mathbf{E}_1^0$, whence corresponding elements in the isomorphism $\psi\{\pi_1(\mathbf{K}_0^n)\} = \pi_1(\mathbf{K}_1^n)$ may be represented by the same product of generators of $\mathfrak{G}_R$. Therefore ϕ is the identical automorphism. Secondly let[28] $D = \sigma$, where σ is a stellar sub-division of some simplicial sub-division of $\mathbf{K}^n$. Then, not only shall $\mathfrak{G}_R$ and $\mathfrak{G}_{R'}$ be identified, but also any natural system $(\mathbf{r}, R)$ for $\mathbf{K}^n$ shall be identified with a natural system for $D(\mathbf{K}^n)$. Finally let $(\mathbf{r}, R)$ be a natural system for $\mathbf{K}^n$, and let $\mathbf{V} \subset \mathbf{K}^1$ be the tree used in defining R. After a suitable sub-division of some simplicial sub-division of $\mathbf{K}^n$ (e.g. s_V^2, using the notation explained in S.S., p. 251) let $D(\mathbf{K}^n)$ consist of shrinking an edge in $\mathbf{V}$ into a point.[29] Then D may be copied by a similar deformation of $\tilde{\mathbf{K}}^n$, and $(\mathbf{r}, R)$ obviously determines a natural system for $D(\mathbf{K}^n)$, which we identify with $(\mathbf{r}, R)$. Thus, by reiterating transformations of this form, $\mathbf{V}$ may be shrunk into a point without altering $(\mathbf{r}, R)$.

Let $F(\mathbf{E}_{0\lambda}^p) \to F(\mathbf{E}_{1\lambda}^p)$ ($\lambda = 1, \cdots, \alpha_p$; $2 \leqq p \leqq n$) be a homotopic deforma-

[28] Cf. S.S., theorem 1.

[29] Cf. S.S., theorem 3.

tion over $\mathbf{K}_0^{p-1}$, where $\mathbf{E}_{01}^p, \cdots, \mathbf{E}_{0\alpha_p}^p$ are the p-cells of a membrane complex $\mathbf{K}_0^n$. Let $\mathbf{K}_1^p = D_p(\mathbf{K}_0^p)(\text{rel. } \mathbf{K}_0^{p-1})$ be a 'formal extension' of the homotopic deformation $F(\mathbf{E}_{0\lambda}^p) \rightarrow F(\mathbf{E}_{1\lambda}^p)$. That is to say

$$\begin{aligned} \mathbf{K}_1^p &= D_p(\mathbf{K}_0^p) \qquad\qquad (\text{rel. } \mathbf{K}_0^{p-1}) \\ &= \mathbf{K}_0^{p-1} + \mathbf{E}_{11}^p + \cdots + \mathbf{E}_{1\alpha_p}^p, \end{aligned}$$

and D_p is given by the construction used in proving S.S., theorem 11, applied to each cell $E_{0\lambda}^p$. We shall call $D_p(\mathbf{K}_0^p)$, or any extension $D_q(\mathbf{K}_0^q)$ $(q = p + 1, \cdots, n)$, of $D_p(\mathbf{K}_0^p)$, a *trivial deformation.* If D_p is a trivial deformation the arguments which lead up to (3.9) remain valid when $q = p$. Also $\mathfrak{G}_R = \mathfrak{G}_{R'}$ and $\phi(x) = x$. Therefore (3.10), with $\bar{r}_{\lambda a}^p = r_{\lambda a}^p$, is valid for $q = p + 1, \cdots, n$. Since $F(\mathbf{E}_{1\lambda}^p)$ is homotopic to $F(\mathbf{E}_{0\lambda}^p)$ it follows that

$$\partial \tilde{\mathbf{E}}_{1\lambda}^p = \partial \tilde{\mathbf{E}}_{0\lambda}^p,$$

where $\tilde{\mathbf{E}}_{i\lambda}^p$ $(i = 0, 1)$ are the cells which appear in (3.9), with $q = p$. Therefore, with the natural choice of basis cells in $\tilde{\mathbf{K}}_1^p$, the matrix $\mathbf{r}^p$ is unaltered by the deformation. Also $\mathbf{r}^q$ is unaltered if $q < p$, since D_p is relative to $\mathbf{K}_0^{p-1}$. The system R may obviously be taken as a natural system of generators and relations for $\pi_1\{D(\mathbf{K}_0^n)\}$, provided we choose a suitable base point on $F(\mathbf{E}_{1\lambda}^2)$ if $p = 2$. Therefore any natural system, $(\mathbf{r}, R)$, for $\mathbf{K}_0^n$, determines a natural system for $D_n(\mathbf{K}_0^n)$, which we identify with $(\mathbf{r}, R)$. We express this by saying that a *natural system,* $(\mathbf{r}, R)$, *for* $\mathbf{K}_0^n$, *is unaltered by a trivial deformation.*

We conclude this section with a theorem which, together with theorems 8 and 9 below, is a measure of the effectiveness of the theory developed here. The theorem may be compared with note III_3, in §6 below.

THEOREM 5. *If* $n = 3$, *or if* $n > 3$ *and* $\pi_s(\mathbf{K}^n) = 0$ *for* $s = 2, \cdots, n - 2$, *then the nucleus of membrane complex* $\mathbf{K}^n$ *is completely determined by one of its natural systems.*

Let $(\mathbf{r}, R)$ be a natural system both for $\mathbf{K}_0^n$ and for $\mathbf{K}_1^n$, and if $n > 3$ let $\pi_s(\mathbf{K}_\rho^n) = 0$ for $\rho = 0, 1$; $s = 2, \cdots, n - 2$. After shrinking a tree in $\mathbf{K}_\rho^1$ into a point, we assume that $\mathbf{K}_\rho^0$ is a single vertex $\mathbf{E}_\rho^0$, and $\mathbf{K}_\rho^1$ a set of circuits which begin and end at $\mathbf{E}_\rho^0$. Since the system R is the same for both complexes it determines a $(1 - 1)$ correspondence between the circuits in $\mathbf{K}_0^1$ and in $\mathbf{K}_1^1$. We identify each circuit in $\mathbf{K}_0^1$ with its image in this correspondence. Since the relations $R_\lambda = 1$ are the same for both complexes we have $\mathbf{K}_1^2 = D_2(\mathbf{K}_0^2)$, where D_2 is a trivial deformation, which we extend throughout $\mathbf{K}_0^n$. Therefore we assume that $\mathbf{K}_0^2 = \mathbf{K}_1^2$. Let $\mathbf{K}_0^p = \mathbf{K}_1^p = \mathbf{K}^p$, say, where $2 \leqq p < n$. Since the matrix $\mathbf{r}^{p+1}$ is the same for $\mathbf{K}_0^{p+1}$ and for $\mathbf{K}_1^{p+1}$ it follows that

$$\partial \tilde{\mathbf{E}}_{0i}^{p+1} \sim \partial \tilde{\mathbf{E}}_{1i}^{p+1} \qquad \text{in} \qquad \tilde{\mathbf{K}}^p,$$

where $\mathbf{E}_{\rho i}^{p+1}$ $(i = 1, \cdots, \alpha_{p+1})$ are the $(p + 1)$-cells of $\mathbf{K}_\rho^{p+1}$ and $\tilde{\mathbf{E}}_{\rho i}^{p+1} \subset \tilde{\mathbf{K}}_\rho^{p+1}$ $(\rho = 0, 1)$ is the chosen basis element which covers $\mathbf{E}_{\rho i}^{p+1}$. Since $\pi_s(\mathbf{K}^p) = \pi_s(\mathbf{K}_\rho^n) = 0$ for $s = 2, \cdots, p - 1$, and since $\pi_s(\mathbf{K}^p) \simeq \pi_s(\tilde{\mathbf{K}}^p)$ if $s > 1$, where $\simeq$ denotes isomorphism, we have $\pi_s(\tilde{\mathbf{K}}^p) = 0$ for $s = 1, \cdots, p - 1$. Therefore

two maps $f_0(S^p), f_1(S^p) \subset \tilde{\mathbf{K}}^p$ are homotopic to each other if[30] $f_0(S^p) \sim f_1(S^p)$. Therefore $F(\tilde{\mathbf{E}}_{0i}^{p+1})$ is homotopic in $\tilde{\mathbf{K}}^p$ to $F(\tilde{\mathbf{E}}_{1i}^{p+1})$ for each $i = 1, \cdots, \alpha_{p+1}$. Therefore $F(\mathbf{E}_{0i}^{p+1})$ is homotopic in $\mathbf{K}^p$ to $F(\mathbf{E}_{1i}^{p+1})$, whence $\mathbf{K}_1^{p+1} = D_{p+1}(\mathbf{K}_0^{p+1})$, where D_{p+1} is a trivial deformation. The deformation D_{p+1} may be extended throughout $\mathbf{K}_0^n$, and the theorem follows from induction on p.

Notice that, if $\mathbf{K}^1 = \mathbf{K}^0$ or if $\mathbf{K}^1$ is a single circuit, then theorem 5 remains valid if $(\mathbf{r}, R)$ is replaced by a set of natural incidence matrices. For in this case R is determined by $\mathbf{r}^2$ and $\mathbf{r}^1$, and even by the first and second integral incidence matrices.

4. We now consider a system $(\mathbf{r}, R)$ from a purely algebraic point of view. We take R to be a particular (finite) system of generators and relations for a symbolic group $\mathfrak{G}_R$, and $\mathbf{r} = (\mathbf{r}^n, \cdots, \mathbf{r}^3)$ to be a set of matrices whose elements are in the group ring $\mathfrak{R}$, of $\mathfrak{G}_R$. The matrix $\mathbf{r}^{p+1}$ is to have α_{p+1} rows and α_p columns ($p = 2, \cdots, n - 1$; $\alpha_p \geqq 0$), where α_2 is the number of relations in the system R, which may be empty in the sense that it has α_2 relations and no generators. Let $\mathbf{r}^2$ and $\mathbf{r}^1$ be the matrices which are calculated from R by the methods explained in §3. Then it follows from (3.7) that $\mathbf{r}^2\mathbf{r}^1 = 0$, and we require that $\mathbf{r}^p\mathbf{r}^{p-1} = 0$ for $p = 3, \cdots, n$. We shall call $\mathbf{r}^n, \cdots, \mathbf{r}^1$ the *incidence matrices* associated with the system $(\mathbf{r}, R)$. We classify such systems by definitions of L- and L^*-equivalence. These are stated in terms of certain elementary transformations of R, together with transformations of the kind considered in §2, which are applicable to the matrices $\mathbf{r}$. These two sets of transformations are not independent, since the transformations of $\mathbf{r}^3$ from the right are governed by the transformations of R, and conversely. This interlocking between the two sets of transformations enables us to avoid a familiar difficulty, which arises in dealing with deformations in a 2-dimensional complex and is absent in the higher dimensionalities.[31]

The elementary transformations of R shall consist of the following transformations:

T_1. *adding a new generator a_0, and a new relation of the form $xa_0^\epsilon y = 1$ ($\epsilon = \pm 1$), where x and y are products of the original generators;*[32]

T_2. *the inverse of T_1;*

T_3. *adding a new relation which is a consequence of the original relations;*

T_4. *the inverse of T_3.*

It is to be understood any generator a_i, together with a relation of the form $xa_i^\epsilon y = 1$, may be discarded by a transformation T_2, provided $a_i^{\pm 1}$ does not occur in x, y or in any of the other relations. If $R' = T_1(R)$ the symbolic group $\mathfrak{G}_{R'}$ is to be regarded as distinct from $\mathfrak{G}_R$, and we associate with T_1 the isomorphism $\phi(\mathfrak{G}_R) = \mathfrak{G}_{R'}$, such that $\phi(a_i) = a_i$, where $a_1, \cdots, a_k$ are the generators of R.

[30] W. Hurewicz, Proc. Kon. Akad. Amsterdam, 38 (1935), 521–8, theorem 1, p. 522.

[31] See, for example, theorem VII in H. Hopf and E. Pannwitz, Math. Annalen, 108 (1933), 433–65.

[32] By a product of the generators we mean a product of the generators and their inverses.

With the transformation $R' = T_2(R)$ we associate the isomorphism $\phi(G_R) = G_{R'}$, such that $\phi(a_i) = a_i$, $\phi(a_0) = (yx)^{-\epsilon}$, where a_0 is the generator and $xa_0^\epsilon y = 1$ is the relation to be discarded. If $R' = T_i(R)$, where $i = 3$ or 4, then $\mathfrak{G}_R = \mathfrak{G}_{R'}$ and ϕ shall be the identical automorphism of $\mathfrak{G}_R$. Thus there is, in each case, a definite isomorphism $\phi(\mathfrak{G}_R) = \mathfrak{G}_{R'}$ associated with the transformation T_i. Therefore a given transformation $T(R) = R'$, where T is the resultant of transformations of the form $T_1, \cdots, T_4$, determines a unique isomorphism $\phi(\mathfrak{G}_R) = \mathfrak{G}_{R'}$. It follows from the standard proof of a familiar theorem[33] that any isomorphism of $\mathfrak{G}_R$ on an isomorphic group, which is given by a finite system of generators and relations, can be expressed as a product of the transformations $T_1, \cdots, T_4$. But, as we shall see, a system R, as part of a system $(\mathbf{r}, R)$, cannot necessarily be transformed into an equivalent system $R' \subset (\mathbf{r}', R')$, since T_4 is to be forbidden except when it is consistent with a corresponding transformation of the matrix $\mathbf{r}^3$.

We now consider the transformations T_i in relation to the matrices $\mathbf{r}^n, \cdots, \mathbf{r}^3$. Let $R' = T_i(R)$ and let $\phi(\mathfrak{R}) = \mathfrak{R}'$ be the isomorphism between the group rings of $\mathfrak{G}_R$ and $\mathfrak{G}_{R'}$, which is determined by the corresponding isomorphism $\phi(\mathfrak{G}_R) = \mathfrak{G}_{R'}$. We first of all replace $\mathbf{r}^p$ by the matrix[34] $\phi(\mathbf{r}^p)$ $(p = 3, \cdots, n)$. Then T_1 shall be accompanied by the addition of a 0^{th} column of zeros to $\phi(\mathbf{r}^3)$. If T_2 is applicable to R, then the i^{th} column of $\mathbf{r}^2$ has $\pm x$ in the λ^{th} row, for some $x \in \mathfrak{G}_R$, and zeros everywhere else, where a_i is the generator and $R_\lambda = 1$ the relation to be discarded. It follows from the relation $\mathbf{r}^3\mathbf{r}^2 = 0$ that the λ^{th} column of $\phi(\mathbf{r}^3)$ consists entirely of zeros, and T_2 is to be accompanied by the removal of this column from $\phi(\mathbf{r}^3)$. Let $R_0 = 1$ be the relation added by T_3, where

$$R_0 = x_1 R_{\lambda_1}^{\epsilon_1} x_1^{-1} \cdots x_q R_{\lambda_q}^{\epsilon_q} x_q^{-1} \qquad (x_i \in \mathfrak{G}_R). \tag{4.1}$$

Then

$$\partial \tilde{\mathbf{E}}_0^2 = \epsilon_1 x_1 \partial \tilde{\mathbf{E}}_{\lambda_1}^2 + \cdots + \epsilon_q x_q \partial \tilde{\mathbf{E}}_{\lambda_q}^2, \tag{4.2}$$

or, after simplification,

$$\partial \tilde{\mathbf{E}}_0^2 = c_1 \partial \tilde{\mathbf{E}}_1^2 + \cdots + c_{\alpha_2} \partial \tilde{\mathbf{E}}_{\alpha_2}^2, \tag{4.3}$$

for certain values of $c_i \in \mathfrak{R}$, where $\partial \tilde{\mathbf{E}}_t^2$ $(t = 0, \cdots, \alpha_2)$ is the linear form (3.4) when R_t is the x in (3.3), and $\tilde{\mathbf{E}}_i^1$ are now to be regarded as undefined basis elements for a modulus with coefficients in $\mathfrak{R}$. The addition of $R_0 = 1$ to the

[33] K. Reidemeister, Einführung in die kombinatorische Topologie, Brunswick (1932), 48.

[34] If $T = T_1$ the same formal expressions may be used to represent $r_{\lambda\alpha}^p$ as $\phi(r_{\lambda\alpha}^p)$, only $r_{\lambda\alpha}^p$ and $\phi(r_{\lambda\alpha}^p)$ must be interpreted as elements of different rings. If $T = T_2$ we first of all substitute yx for $a_0^{-\epsilon}$, whenever the latter appears in given formal expressions for $r_{\lambda\alpha}^p$. This does not alter the element $r_{\lambda\alpha}^p \in \mathfrak{R}$, only its formal expression as a polynomial in the generators. The transformation $r_{\lambda\alpha}^p \to \phi(r_{\lambda\alpha}^p)$ may then be defined as when $T = T_1$. If $T = T_3$ or T_4 we have $\phi(\mathbf{r}^p) = \mathbf{r}^p$.

relations of R is to be accompanied by an expanion of the form

$$\mathbf{r}^3 \to \left\| \begin{matrix} \epsilon x & -\epsilon x\mathbf{c} \\ 0 & \mathbf{r}^3 \end{matrix} \right\|, \tag{4.4}$$

where $\epsilon = \pm 1$, $x \in \mathfrak{G}_R$ is arbitrary and $-\epsilon x\mathbf{c}$ stands for the row $\| -\epsilon xc_1, \cdots, -\epsilon xc_{\alpha_2} \|$ together with the corresponding secondary expansion of $\mathbf{r}^4$. Conversely, let (4.4) be a given expansion, in which x and c_i, and hence $-\epsilon xc_i$, are arbitrary, and let (4.3) be expressed in the form (4.2). Then we accompany the expansion (4.4) by adding the new relation $R_0 = 1$, where R_0 is of the form (4.1), with the factors $x_i R_{\lambda_i}^{\epsilon_i} x_i^{-1}$ arranged in an arbitrary order. A transformation of the form T_4, in which $R_\lambda = 1$ is the relation to be discarded, shall only be applied if the λ^{th} column of $\mathbf{r}^3$ has $\pm x$ ($x \in \mathfrak{G}_R$) in the i^{th} row, for an arbitrary $i = 1, \cdots, \alpha_3$, and zeros everywhere else. In this case it is to imply the contraction of $\mathbf{r}^3$, in which the i^{th} row and the λ^{th} column are discarded, and the corresponding contraction of $\mathbf{r}^4$. Conversely, such a contraction of $\mathbf{r}^3$ is only to be applied when $R_\lambda = 1$ is a consequence of the other relations. In this case it is to be accompanied by the transformation T_4, in which the relation $R_\lambda = 1$ is discarded. It is now to be understood that a transformation $T_i(\mathbf{r}, R)$ consists of the transformation $R \to T_i(R)$, together with the transformation $\mathbf{r}^p \to T_i(\mathbf{r}^p)$, where $T_i(\mathbf{r}^4)$, $T_i(\mathbf{r}^3)$ are the matrices into which $\mathbf{r}^4$, $\mathbf{r}^3$ are transformed by the above rules, and $T_i(\mathbf{r}^p) = \phi(\mathbf{r}^p)$ if $p > 4$, or if $p = 4$ and $i = 1$ or 2.

In addition to the transformations T_i we allow the matrices $\mathbf{r}^n, \cdots, \mathbf{r}^4$ to be transformed as in §2. That is to say, we allow arbitrary expansions and contractions of $\mathbf{r}^p$ ($p = 4, \cdots, n$), which must be accompanied by the appropriate secondary expansions and contractions of $\mathbf{r}^3$ if $p = 4$, and transformations of the form

$$\mathbf{r}^{p+1} \to \mathbf{r}^{p+1}\mathbf{a}_p^{-1}, \qquad \mathbf{r}^p \to \mathbf{a}_p\mathbf{r}^p \qquad (p = 3, \cdots, n),$$

where $\mathbf{a}_p$ is any regular, or elementary, matrix of degree α_p, according to the kind of equivalence considered. Two systems $(\mathbf{r}, R)$ and $(\mathbf{r}', R')$, will be described as *X-equivalent* ($X = L$ or L^*) if, and only if, they are equivalent under the combined set of transformations, in which $\mathbf{a}_p$ may be any regular matrix, of the appropriate degree, if $X = L^*$, and is to be an elementary matrix if $X = L$. More precisely, we shall describe $(\mathbf{r}, R)$ and $(\mathbf{r}', R')$ as *X-equivalent under the isomorphism* ϕ, where $\phi(\mathfrak{G}_R) = \mathfrak{G}_{R'}$ is the isomorphism induced by the transformation $T(R) = R'$. It is important to notice that ϕ depends on the actual sequence of the transformations T_i which appear in the transformation T, not only on the final result. For example, ϕ need not be the identical automorphism if $\mathfrak{G}_R = \mathfrak{G}_{R'}$, or even if $R' = R$. Clearly theorems 1 and 2 are valid when restated in terms of equivalence between systems $(\mathbf{r}, R)$ and $(\mathbf{r}', R')$. Let us describe as an *elementary expansion* of $(\mathbf{r}, R)$ either a transformation of the form T_1, or T_3, or an expansion of the form (2.5), where $p > 3$, and let us describe the inverse of an elementary expansion as an *elementary contraction*. It follows from the proof of lemma 4 that $(\mathbf{r}, R)$ can be transformed

into any L-equivalent system by a series of elementary expansions and contractions, provided we allow primary expansions of the empty matrix $\mathbf{r}^{n+1}$.

Let κ_ν denote the ν^{th} column of $\mathbf{r}^3$, let x be a given element of $\mathfrak{G}_R$, and let λ, μ $(\lambda \neq \mu)$ be particular values of $\nu = 1, \cdots, \alpha_2$. I say that the transformation

$$R_\lambda \to R'_\lambda = xR^\epsilon_\lambda x^{-1}R_\mu \qquad (\epsilon = \pm 1), \tag{4.5}$$

accompanied by $\kappa_\lambda \to \epsilon\kappa_\lambda x^{-1}$, $\kappa_\mu \to \kappa_\mu - \epsilon\kappa_\lambda x^{-1}$, is the resultant of a transformation T_3, followed by elementary transformations of $\mathbf{r}^4$, $\mathbf{r}^3$, whose resultant leaves $\mathbf{r}^4$ unaltered, followed by T_4. For we first add the relation $R'_\lambda = 1 (R'_\lambda = R_0)$ to R and the row r^3_{0i} $(i = 0, \cdots, \alpha_2)$ to $\mathbf{r}^3$, where

$$r^3_{00} = 1, \qquad r^3_{0\lambda} = -\epsilon x, \qquad r^3_{0\mu} = -1, \qquad r^3_{0\nu} = 0 \qquad (\nu \neq \lambda, \mu),$$

and the columns $r^3_{a0} = \delta_{a0}$, $r^4_{k0} = 0$ $(a = 0, \cdots, \alpha_3;\ k = 1, \cdots, \alpha_4)$ to $\mathbf{r}^3$ and $\mathbf{r}^4$. We then clear away the elements of $\mathbf{r}^3$, which stand under $r^3_{0\lambda}$, by the transformations $\rho_t \to \rho_t + \epsilon r^3_{t\lambda}x^{-1}\rho_0$ $(t = 1, \cdots, \alpha_3)$, where ρ_t stands for the t^{th} row of $\mathbf{r}^3$. As in lemma 4, it follows from the relation $\mathbf{r}^4\mathbf{r}^3 = 0$ that the corresponding transformation of $\| 0, \mathbf{r}^4 \|$ leaves the latter unchanged. We now discard the relation $R_\lambda = 1$, and the appropriate rows and columns of the expanded matrices $\mathbf{r}^4$, $\mathbf{r}^3$, and rearrange the notation so that $R'_\lambda = 1$ appears as the λ^{th} relation. The final result is the transformation (4.5). Notice that the transformation $R_\lambda \to xR^\epsilon_\lambda x^{-1}$ is the resultant of transformations of the form (4.5), namely

$$R_\lambda \to xR^{-\epsilon}_\lambda x^{-1}R_\mu \to R_\mu(xR^{-\epsilon}_\lambda x^{-1}R_\mu)^{-1}R^{-1}_\mu R_\mu = xR^\epsilon_\lambda x^{-1}.$$

Let $(\mathbf{r}, R)$ and $(\mathbf{r}', R')$ be two given systems, and let $\mathbf{r}^2$, $\mathbf{r}^1$ and $\mathbf{r}'^2$, $\mathbf{r}'^1$ be the second and first two incidence matrices determined by R and R'.

THEOREM 6. *If the systems* $(\mathbf{r}, R)$ *and* $(\mathbf{r}', R')$ *are X-equivalent* $(X = L$ *or* $L^*)$ *under a given isomorphism* ϕ, *then the sets of incidence matrices* $\mathbf{r}^n, \cdots, \mathbf{r}^1$ *and* $\mathbf{r}'^n, \cdots, \mathbf{r}'^1$ *are X-equivalent under* ϕ.

This relation between $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{r}'^n, \cdots, \mathbf{r}'^1$ is obviously transitive, and if $R' = R$ and R is left unaltered throughout the passage from $(\mathbf{r}, R)$ to $(\mathbf{r}', R')$ there is nothing to prove. Therefore the theorem will follow from an inductive argument when we have proved it in case T is a single transformation T_i. If $i = 3$ or 4 the theorem is implicit in the definition of $T_i(\mathbf{r}, R)$. On replacing ϕ by ϕ^{-1} the equivalence relation between $\mathbf{r}^n, \cdots, \mathbf{r}^1$ and $\mathbf{r}'^n, \cdots, \mathbf{r}'^1$ is seen to be symmetric. Therefore it is sufficient to consider the case where $i = 1$. If a relation $R_0 = xa^\epsilon_0 y = 1$ is replaced by $zR^{\pm 1}_0 z^{-1} = 1$ $(z \in \mathfrak{G}_R)$, the 0^{th} row of $\phi(\mathbf{r}^2)$ is multiplied by $\pm z$. Therefore we may assume that $R_0 = 0_0 x$, for some x which is a product of the existing generators. Then the 0^{th} row of $\mathbf{r}'^2$ is $\| 1, a_0 r_1, \cdots, a_0 r_k \|$, where r_i means the same as it does in (3.4) when x is given by (3.3). The i^{th} row of $\mathbf{r}'^1$ is $a_i - 1$ $(i = 0, \cdots, k)$, and since $x = a_0^{-1}$, in consequence of $R_0 = 1$, it follows from (3.7) that

$$\begin{aligned} a_0 - 1 &= a_0(1 - x) \\ &= -\sum_{i=1}^{k} a_0 r_i(a_i - 1), \end{aligned}$$

whence $\mathbf{r}'^2$, $\mathbf{r}'^1$ are obtained from $\phi(\mathbf{r}^2)$, $\phi(\mathbf{r}^1)$ by an expansion of the form (2.5). The appropriate column of zeros is added to $\phi(\mathbf{r}^3)$, and the theorem is established.

We now prove a theorem which shows how the generators of R can be manipulated by what may be called the 'free group calculus.' Let the transformation

$$a_i \to f_i(a'_1, \cdots, a'_k) = f_i(a') \qquad (i = 1, \cdots, k)$$

determine an isomorphism of $\mathfrak{F}_k$ on the free group $\mathfrak{F}'_k$, which is freely generated by $a'_1, \cdots, a'_k$. Some or all of the generators $a'_1, \cdots, a'_k$ may coincide with some of $a_1^{\pm 1}, \cdots, a_k^{\pm 1}$, so that the possibility $\mathfrak{F}'_k = \mathfrak{F}_k$ is not excluded. Let

$$R'_\lambda = R_\lambda\{f_1(a'), \cdots, f_k(a')\} \qquad (\lambda = 1, \cdots, \alpha_2),$$

where $R_\lambda(a_1, \cdots, a_k) = 1$ are the relations of the system R, and let R' be the system which consists of the generators $a'_1, \cdots, a'_k$, subject to the relations $R'_\lambda = 1$. Since the transformation given by $a_i \to f_i(a')$ is an isomorphism of $\mathfrak{F}_k$ on $\mathfrak{F}'_k$, the group $\mathfrak{G}_{R'}$ is also generated by $f_1(a'), \cdots, f_k(a')$, and $a_i \to f_i(a')$ determines an isomorphism $\phi(\mathfrak{G}_R) = \mathfrak{G}_{R'}$, in which $\phi(a_i) = f_i(a')$. Let $\mathbf{r}'^s = \phi(\mathbf{r}^s)$ $(s = 3, \cdots, n)$, where $\mathbf{r}^n, \cdots, \mathbf{r}^3$ are the matrices of a given system $(\mathbf{r}, R)$. Then we have the theorem:

THEOREM 7. *The system* $(\mathbf{r}, R)$ *is L-equivalent under the isomorphism* ϕ *to the system* $(\mathbf{r}', R')$.

First assume that $a'_i \neq a_j^{\pm 1}$ for any values of i, j and that either

1. $f_p(a') = a_p'^{\epsilon}$, $\quad f_i(a') = a'_i (\epsilon = \pm 1, i \neq p)$, or
2. $f_p(a') = a'_p a'_q$, $f_i(a') = a'_i$ $(q, i \neq p)$.

Let $a'_i \to g_i(a)$ be the inverse of the isomorphism $a_i \to f_i(a')$, so that $g_p(a) = a_p^\epsilon$, $g_i(a) = a_i$, in the first case, and $g_p(a) = a_p a_q^{-1}$, $g_i(a) = a_i$ in the second. Let $T(R) = \bar{R}$ be the expansion of R in which $a'_1, \cdots, a'_k$ are added as new generators, together with the relations $a_i'^{-1} g_i(a) = 1$, let $\phi_1(\mathfrak{G}_R) = \mathfrak{G}_{\bar{R}}$, given by $\phi_1(a_i) = a_i$, be the corresponding isomorphism of $\mathfrak{G}_R$, and let $\bar{\mathbf{r}}^3 = T\{\phi_1(\mathbf{r}^3)\}$. Then $\bar{\mathbf{r}}^3 = \| \phi_1(\mathbf{r}^3), 0 \|$, where 0 indicates the columns which correspond to the new relations. The latter can obviously be transformed into the relations $a_i^{-1} f_i(a') = 1$ by transformations of the form (4.5), and these transformations leave $\bar{\mathbf{r}}^3$ unaltered, since the new columns consist entirely of zeros. We now transform $R_\lambda(a_1, \cdots, a_k) = 1$ into $R'_\lambda(a'_1, \cdots, a'_k) = 1$, for each $\lambda = 1, \cdots, \alpha_2$, by substituting $f_i(a')$ for a_i throughout the products R_λ. This operation is the resultant of transformations of the form

$$xa_iy' \to xa_iy'y'^{-1}a_i^{-1}f_iy' \qquad \text{or} \qquad xa_i^{-1}y' \to xf_i^{-1}a_ix^{-1}xa_i^{-1}y',$$

where $f_i = f_i(a')$, $xa_i^\epsilon y = R_\lambda$, for some $\lambda = 1, \cdots, \alpha_2$, and y' is the result of substituting $f_i(a')$ for a_i throughout the product y. Therefore it follows without difficulty that the transformation $R_\lambda \to R'_\lambda$ is the resultant of transformations of the form (4.5). Moreover the columns of $\mathbf{r}^3$, which correspond to the original relations $R_\lambda = 1$, are unaltered by these transformations.[35] Finally we discard

[35] As explained above, in the definition of $T_2(\mathbf{r}^3)$, it follows from the condition $\mathbf{r}^4\mathbf{r}^3 = 0$, and the form of the relations $a_i^{-1}f_i(a') = 1$, that the remaining (zero) columns also are ultimately left unaltered.

the generators $a_i, \cdots, a_k$ and the relations $a_i^{-1}f_i(a') = 1$ by transformations of the form T_2. Then $\phi_1(\mathbf{r}^p)$ $(p = 4, \cdots, n)$ is transformed into $\phi_2(\bar{\mathbf{r}}^p) = \phi_2\phi_1(\mathbf{r}^p)$, where $\phi_2(a_i') = a_i'$, $\phi_2(a_i) = f_i(a')$, and $\phi_2\phi_1 = \phi$, since $\phi_2\phi_1(a_i) = \phi_2(a_i) = f_i(a')$. Similarly $\mathbf{r}^3$ contracts into $\phi(\mathbf{r}^3) = \mathbf{r}'^3$ on discarding the (zero) columns, which correspond to the relations $a_i^{-1}f_i(a') = 1$. Therefore $(\mathbf{r}, R)$ is L-equivalent to $(\mathbf{r}', R')$ under the isomorphism ϕ.

What we have already proved, with $f_i(a') = a_i'$, allows us to 're-write' a_i as a_i' if $a_i' \neq a_j^{\pm 1}$, for any i, j and even if the sets a_i' and $a_j^{\pm 1}$ overlap. For we can first rewrite a_i as b_i and then re-write b_i as a_i', where $b_i^{\pm 1} \neq a_j$ or a_j'. Therefore, in completing the proof, we may assume that $a_i' = a_i$, in which case $a_i \rightarrow f_i(a)$ is an automorphism of $\mathfrak{F}_k$. Moreover we may re-write a_i' as a_i in what we have already proved. Therefore the theorem follows from the fact that the automorphisms of $\mathfrak{F}_k$ are generated by automorphisms of the form $a_i \rightarrow f_i(a)$, where[36] $f_p(a) = a_p^{-1}$ or a_pa_q, $f_i(a) = a_i$ $(q, i \neq p)$.

THEOREM 8. *If* $\mathbf{K}_1^n$ *and* $\mathbf{K}_2^n$ *have the same nucleus, then any natural systems for* $\mathbf{K}_1^n$ *and* $\mathbf{K}_2^n$ *are L-equivalent. If* $\mathbf{K}_1^n$ *and* $\mathbf{K}_2^n$ *are of the same homotopy type, then their natural systems are* L^**-equivalent.*

This will obviously follow from the theorems 1 and 2, the arguments in Q.J., and S.S., theorem 13, when we have shown that two natural systems for the same complex $\mathbf{K}^n$, are L-equivalent. Let R and R' be two natural systems of generators and relations for $\pi_1(\mathbf{K}^n)$, and let $\phi(\mathfrak{G}_R) = \mathfrak{G}_{R'}$ be the isomorphism such that a given element in $\pi_1(\mathbf{K}^n)$ is represented by $x \in \mathfrak{G}_R$ and $\phi(x) \in \mathfrak{G}_{R'}$ After a change of basis, if necessary, we may obviously assume that $\mathbf{r}'^p = \phi(\mathbf{r}^p)$ $(p = 3, \cdots, n)$, where $\mathbf{r}^n, \cdots, \mathbf{r}^3$ and $\mathbf{r}'^n, \cdots, \mathbf{r}'^3$ are the matrices in natural systems, $(\mathbf{r}, R)$ and $(\mathbf{r}', R')$, for $\mathbf{K}^n$. Let $\mathbf{V}, \mathbf{V}' \subset \mathbf{K}^1$ be the trees in terms of which R and R' are defined. Then we may also assume that $\mathbf{V}' = \mathbf{V} - \mathbf{E}_1^1 + \mathbf{E}_2^1$, where $\mathbf{E}_1^1 \subset V$, $\mathbf{E}_2^1 \subset \mathbf{K}^1 - V$, and $\mathbf{E}_1^1$ separates the extremities of $\mathbf{E}_2^1$. For one can interchange any two trees, which are sub-complexes of $\mathbf{K}^1$ and contain $\mathbf{K}^0$, by reiterating transformations of the form $\mathbf{V} \rightarrow \mathbf{V}'$, where $\mathbf{V}$ and $\mathbf{V}'$ are thus related.[37]

On these assumptions let us, without altering the notation, shrink into a point each of the two trees into which $\mathbf{V}$ is separated by $\mathbf{E}_1^1$. Then $\mathbf{K}^1$ will consist of two vertices $\mathbf{E}_1^0$, $\mathbf{E}_2^0$, which are joined by certain 1-cells $\mathbf{E}_1^1, \mathbf{E}_2^1, \cdots, \mathbf{E}_p^1$, and of circuits $\mathbf{E}_{h1}^1, \cdots, \mathbf{E}_{hm_h}^1$ $(h = 1, 2)$ beginning and ending at $\mathbf{E}_h^0$. By theorem 7 we may orient $\mathbf{E}_i^1$ as we please, and we take $\mathbf{E}_1^0$ to be its first point. Let us denote $\mathbf{E}_i^1$ by s_i $(i = 1, \cdots, p)$ and $\mathbf{E}_{h\rho}^1$ by $\sigma_{h\rho}$. Then, as in (3.3), any segment on $\mathbf{K}'$ may be represented by a word

$$t_1^{\epsilon_1} \cdots t_q^{\epsilon_q}, \tag{4.6}$$

where $t_\lambda = s_{i_\lambda}$, or $\sigma_{h_\lambda \rho_\lambda}$. Let us now write $\sigma_{2\rho} = \sigma_\rho$. I say that, if (4.6) represents a circuit which begins and ends at $\mathbf{E}_1^0$, then, regarded as an element

[36] J. Nielsen, Math. Annalen, 79 (1919), 269-72.

[37] Cf. p. 105 of K. Reidemeister, Abh. Hamb. Sem., 11 (1935), 102-9.

in the free group, which is freely generated by s_i, σ_ρ, σ_{1a}, it is unaltered when we replace s_j by $s_j s_i^{-1}$ and σ_ρ by $s_i\sigma_\rho s_i^{-1}$, for each $j = 1, \cdots, p$, $\rho = 1, \cdots, m_2$, and any specified value of i. For each letter of the form s_j or $\sigma_\rho^{\pm 1}$ represents a segment which terminates at $\mathbf{E}_2^0$. Since the segment (4.6) terminates at $\mathbf{E}_1^0$ such a letter is followed by s_l^{-1} or by $\sigma_\tau^{\pm 1}$ for some value of l or ρ. Similarly s_j^{-1} or $\sigma_\rho^{\pm 1}$ is preceded by s_l or $\sigma_\tau^{\pm 1}$, since (4.6) begins at $\mathbf{E}_1^0$. Therefore, taking $\mathbf{E}_1^0$ as base point, we have $F(\mathbf{E}_\lambda^2) = R_\lambda(s_j, \sigma_\rho, \sigma_{1a})$, where

$$R_\lambda(s_j, \sigma_\rho, \sigma_{1a}) = R_\lambda(s_j s_i^{-1}, s_i\sigma_\rho s_i^{-1}, \sigma_{1a}),$$

in which the equality refers to the free group generated by s_i, σ_ρ, σ_{1a}. Taking $i = 1$ we have

$$F(\mathbf{E}_\lambda^2) = R_\lambda(1, a_j, b_\rho, \sigma_{1a}) \qquad (j = 2, \cdots, p), \tag{4.7}$$

where $a_j = s_j s_1^{-1}$, $b_\rho = s_1\sigma_\rho s_1^{-1}$, and the system R, associated with $\mathbf{V}$, consists of the generators $a_2, \cdots, a_p$, b_ρ, σ_{1a}, subject to the relations $R_\lambda = 1$. Similarly the system R', associated with $\mathbf{V}'$, is generated by a_1', a_3', $\cdots$, a_p', b_ρ', σ_{1a}, subject to the relations $R_\lambda' = 1$, where $a_t' = s_t s_2^{-1}$ $(t = 1, 3, \cdots, p)$, $b_\rho' = s_2\sigma_\rho s_2^{-1}$ and

$$R_\lambda' = R_\lambda(a_1', 1, a_\beta', b_\rho', \sigma_{1a}) \qquad (\beta = 3, \cdots, p). \tag{4.8}$$

On writing $s_2 = 1$ in the equation

$$R_\lambda(1, s_2 s_1^{-1}, s_\beta s_1^{-1}, s_1\sigma_\rho s_1^{-1}, \sigma_{1a}) = R_\lambda(s_1, s_2, s_\beta, \sigma_\rho, \sigma_{1a}),$$

we have

$$R_\lambda(1, s_1^{-1}, s_\beta s_1^{-1}, s_1\sigma_\rho s_1^{-1}, \sigma_{1a}) = R_\lambda(s_1, 1, s_\beta, \sigma_\rho, \sigma_{1a}). \tag{4.9}$$

On comparing (4.7) with (4.8), it follows from (4.9) that R' is the image of R in the transformation given by

$$a_2 \to a_1'^{-1}, \qquad a_\beta \to a_\beta' a_1'^{-1}, \qquad b_\rho \to a_1' b_\rho' a_1'^{-1}, \qquad \sigma_{1a} \to \sigma_{1a}. \tag{4.10}$$

The product (4.6) is transformed into an element $x \in \mathfrak{G}_R$ if we write $s_j = a_j s_1$, $\sigma_\rho = s_1^{-1} b_\rho s_1$ and cancel all the factors $s_1^{\pm 1}$, and into $\phi_1(x) \in \mathfrak{G}_{R'}$ if we write $s_j = a_j' s_2$, $\sigma_\rho = s_2^{-1} b_\rho' s_2$ and cancel $s_2^{\pm 1}$, where $\phi_1(\mathfrak{G}_R) = \mathfrak{G}_{R'}$ is the isomorphism given by (4.10). Therefore ϕ_1 is the isomorphism ϕ, and since $\mathbf{r}'^p = \phi(\mathbf{r}^p)$ the theorem follows from theorem 7.

As in theorem 3 it is clear that the isomorphism $\psi\{\pi_1(\mathbf{K}_1^n)\} = \pi_1(\mathbf{K}_2^n)$, which is expressed algebraically by the isomorphism $\phi(\mathfrak{G}_R) = \mathfrak{G}_{R'}$, may be taken from the class which is determined by some homotopy isomorphism $f(\mathbf{K}_1^n) \subset \mathbf{K}_2^n$. The distinction between ϕ and ψ is analogous to one which is familiar in coordinate geometry, where an algebraic transformation may represent either a transformation of coordinates or a point transformation, according to the context.

We now prove an analogous theorem to theorem 8, in which the parts played by the algebraic and the geometric equivalence are interchanged. In general the analogy is not complete. For two complexes with the same natural system

need not even be of the same homotopy type.[38] Moreover, there is no reason to suppose that every system $(\mathbf{r}, R)$ can be realized geometrically as a natural system of a complex. The theorem is

THEOREM 9. *Any system $(\mathbf{r}', R')$, which is X-equivalent ($X = L$ or L^*) to a natural system, $(\mathbf{r}, R)$, for a given complex $\mathbf{K}^n$, is a natural system for some complex of the same homotopy type as $\mathbf{K}^n$, and for some complex with the same nucleus of $X = L$.*

First let $X = L$. Then $(\mathbf{r}, R)$ can be transformed into $(\mathbf{r}', R')$ by a sequence of elementary expansions and contractions, according to lemma 4. Therefore the theorem will follow inductively, for $X = L$, when we have proved it in case $(\mathbf{r}', R')$ is obtained from $(\mathbf{r}, R)$ by a single elementary expansion or contraction.

Let $(\mathbf{r}', R')$ be obtained from $(\mathbf{r}, R)$ by an elementary expansion. First let $(\mathbf{r}', R') = T_1(\mathbf{r}, R)$, and let a_0 be the new generator and $xa_0^\epsilon y = 1$ the new relation. Let $s, t \subset \mathbf{K}^1$ be oriented circuits, beginning and ending at the base point $\mathbf{E}^0 \in \mathbf{K}^0$, which are represented by x and y respectively. Let $\mathbf{E}_0^1$ be a new 1-cell, beginning and ending at $\mathbf{E}^0$ which shall correspond to the new generator a_0. Let $\mathbf{E}_0^2$ be a new 2-cell such that $F(\mathbf{E}_0^2) \subset \mathbf{K}^1 + \mathbf{E}_0^1$ is given by $xa_0^\epsilon y$, with a suitable base point. Then $(\mathbf{r}', R')$ is a natural system for $\mathbf{K}_1^n = \mathbf{K}^n + \mathbf{E}_0^2$, and assuming that $F(\mathbf{E}_0^2)$ is simple in $\mathbf{E}_0^1$, as we obviously may, $\mathbf{K}^n$ expands into $\mathbf{K}_1^n$. Secondly let $(\mathbf{r}', R') = T_3(\mathbf{r}, R)$ and let $R_0 = 1$ be the new relation, where R_0 is given by (4.1). With the same notation as in §3, there is obviously a singular 2-cell $\tilde{\mathbf{G}}^2 \subset \tilde{\mathbf{K}}^2$, whose boundary, $F(\tilde{\mathbf{G}}^2) \subset \tilde{\mathbf{K}}^1$, is represented by the product R_0, such that the relative 2-cycle, mod. $\tilde{\mathbf{K}}^1$, determined by $\tilde{\mathbf{G}}^2$ is

$$\epsilon_1 x_1 \tilde{\mathbf{E}}_{\lambda_1}^2 + \cdots + \epsilon_q x_q \tilde{\mathbf{E}}_{\lambda_q}^2 = c_1 \tilde{\mathbf{E}}_1^2 + \cdots + c_{\alpha_2} \tilde{\mathbf{E}}_{\alpha_2}^2 \qquad (c_i \in \mathfrak{R}).$$

Let $\mathbf{E}_0^2$ be a new membrane whose boundary is given by $F(\mathbf{E}_0^2) = u\{F(\tilde{\mathbf{G}}^2)\} \subset \mathbf{K}^1$, let $\mathbf{G}^2 = u(\tilde{\mathbf{G}}^2) \subset \mathbf{K}^2$ and let $\mathbf{E}_0^3$ be a new membrane bounded by the spherical map $\epsilon(\mathbf{E}_0^2 - \mathbf{G}^2)$ $(\epsilon = \pm 1)$. Then $F(\mathbf{E}_0^3)$ is simple in $\mathbf{E}_0^2$, whence $\mathbf{K}^n$ expands into $\mathbf{K}_1^n = \mathbf{K}^n + \mathbf{E}_0^3$. Assuming that $\tilde{\mathbf{K}}^n$ is imbedded in $\tilde{\mathbf{K}}_1^n$, let $\tilde{\mathbf{E}}_0^2 \in u^{-1}(\mathbf{E}_0^2) \subset \tilde{\mathbf{K}}_1^2$, and $\tilde{\mathbf{E}}_0^3 \in u^{-1}(\mathbf{E}_0^3) \subset \tilde{\mathbf{K}}_1^3$ be chosen so that $F(\tilde{\mathbf{E}}_0^2) = F(\tilde{\mathbf{G}}^2)$ and $F(\tilde{\mathbf{E}}_0^3) = \epsilon x(\tilde{\mathbf{E}}_0^2 - \tilde{\mathbf{G}}^2)$, where

$$\mathbf{r}'^3 = \left\| \begin{matrix} \epsilon x & -\epsilon x \mathbf{c} \\ 0 & \mathbf{r}^3 \end{matrix} \right\|,$$

as in (4.4). Then

$$\partial \tilde{\mathbf{E}}_0^3 = \epsilon x \tilde{\mathbf{E}}_0^2 - \epsilon x(c_1 \tilde{\mathbf{E}}_1^2 + \cdots + c_{\alpha_2} \tilde{\mathbf{E}}_{\alpha_2}^2),$$

whence $(\mathbf{r}', R')$ is a natural system for $\mathbf{K}_1^n$. An elementary expansion of the form (2.5) with $p > 3$, can be copied in the same way.

Let $(\mathbf{r}', R') = T_2(\mathbf{r}, R)$, and let a_0 be the generator and $xa_0^\epsilon y = 1$ the relation to be removed. Since $a_0^{\pm 1}$ does not appear in any other relation we may assume,

[38] Cf. note III_3.

after a trivial deformation of $\mathbf{K}^2$ if necessary, that $F(\mathbf{E}_i^2) \subset \mathbf{K}_1^1$ $(i = 1, \cdots, \alpha_2)$, where $\mathbf{E}_0^1, \cdots, \mathbf{E}_{\alpha_1}^1$ are the 1-cells of $\mathbf{K}^n$, and

$$\mathbf{K}_1^1 = \mathbf{K}^0 + \mathbf{E}_1^1 + \cdots + \mathbf{E}_{\alpha_1}^1 .$$

We also assume that $F(\mathbf{E}_0^2)$ is simple in $\mathbf{E}_0^1$, where $\mathbf{E}_0^1$ and $\mathbf{E}_0^2$ are the cells corresponding to a_0 and $xa_0^{\epsilon}y = 1$. Then $\mathbf{K}^2$ contracts into $\mathbf{K}_1^2$, where

$$\mathbf{K}_1^2 = \mathbf{K}_1^1 + \mathbf{E}_1^2 + \cdots + \mathbf{E}_{\alpha_2}^2 .$$

It follows from an obvious induction that $\mathbf{K}^n$ may be transformed by a series of trivial deformations into a complex of the form $\mathbf{K}_1^n + \mathbf{E}_0^2$, where neither $\mathbf{E}_0^2$ nor $\mathbf{E}_0^1$ has an inner point in common with $\mathbf{K}_1^n$. In other words we push the remaining cells in $\mathbf{K}^n$ off the interiors of $\mathbf{E}_0^1$ and $\mathbf{E}_0^2$. Then $\mathbf{K}_1^n + \mathbf{E}_0^2$ contracts into $\mathbf{K}_1^n$, and $(\mathbf{r}', R')$ is a natural system for $\mathbf{K}_1^n$.

Let $(\mathbf{r}, R) \to (\mathbf{r}', R')$ be due to a contraction of $\mathbf{r}^p$, for some $p > 3$, in which the 0^{th} row and column are removed. Then $\tilde{\mathbf{E}}_0^{p-1}$ appears in $\partial\tilde{\mathbf{E}}_0^p$ with the coefficient $\pm x$, for some $x \in \mathfrak{G}_R$, and with the coefficient zero in $\partial\tilde{\mathbf{E}}_i^p$ $(i = 1, \cdots, \alpha_p)$. Let $\mathbf{A}^{p-1} \in \mathbf{E}_0^{p-1}$ be an open simplex, whose closure is internal to $\mathbf{E}_0^{p-1}$, and let $y\tilde{\mathbf{A}}^{p-1}$ be its image in $y\tilde{\mathbf{E}}_0^{p-1}$, for each $y \in \mathfrak{G}_R$. Then it follows from a standard theorem [39] that $F(\tilde{\mathbf{E}}_i^p)$ $(i = 1, \cdots, \alpha_p)$ is homotopic in $\tilde{\mathbf{K}}^{p-1}$ to a map in which $y\tilde{\mathbf{A}}^{p-1}$ is uncovered for each $y \in \mathfrak{G}_R$. Also $F(\tilde{\mathbf{E}}_0^p)$ is homotopic to a map which is simple in $x\tilde{\mathbf{A}}^{p-1}$ and in which $y\tilde{\mathbf{A}}^{p-1}$ is uncovered if $y \neq x$. It follows that $F(\mathbf{E}_0^p)$ is homotopic in $\mathbf{K}^{p-1}$ to a map which is simple in $\mathbf{E}_0^{p-1}$, and that $F(\mathbf{E}_i^p)$ $(i = 1, \cdots, \alpha_p)$ is homotopic to a map in which $\mathbf{A}^{p-1}$ is uncovered, and hence to a map in

$$\mathbf{K}_1^{p-1} = \mathbf{K}^{p-2} + \mathbf{E}_1^{p-1} + \cdots + \mathbf{E}_{\alpha_{p-1}}^{p-1} .$$

The same is true if $p = 3$ and $(\mathbf{r}', R') = T_4(\mathbf{r}, R)$. For since the relation $R_0 = 1$, which is to be discarded, is redundant, it follows that $\tilde{\mathbf{K}}^2 - u^{-1}(\mathbf{A}^2)$ is simply connected, where $\mathbf{A}^2 \in \mathbf{E}_0^2$. Therefore the above argument, which fails, in general, unless $p - 1 > 2$, is valid here if $p - 1 = 2$. In either case it now follows from the argument used where $(\mathbf{r}', R') = T_2(\mathbf{r}, R)$, that $(\mathbf{r}', R')$ is a natural system for a complex $\mathbf{K}_1^n$, which is obtained from $\mathbf{K}^n$ by a series of trivial deformations, followed by a contraction. Therefore the theorem is established in case $X = L$.

In proving theorem 17 in S.S. it was shown that two n-dimensional complexes are of the same homotopy type if they have the same $(n + 1)$-group. Therefore the theorem will follow, in case $X = L^*$, when we have shown that (x', R') is a natural system for an (n-dimensional) complex with the same $(n + 1)$-group as $\mathbf{K}^n$. Let empty matrices $\mathbf{r}^p$, $\mathbf{r}'^p$ $(p = n + 1, n + 2)$, with α_p, α'_p, rows and α_{p-1}, α'_{p-1} columns $(\alpha_{n+2} = \alpha'_{n+2} \geqq 0, \alpha_{n+1} = \alpha'_{n+1} = 0, \alpha_n, \alpha'_n =$ number of

[39] Cf. P. Alexandroff and H. Hopf, (loc. cit.), p. 503, or pp. 99–100 or S. Lefschetz, Fund. Math., 27 (1936), 94–115.

rows in $\mathbf{r}^n$, $\mathbf{r}'^n$), be included in the systems $(\mathbf{r}, R)$ and $(\mathbf{r}', R')$. The augmented system $(\mathbf{r}, R)$ is a natural system for any complex of the form

$$\mathbf{K}^{n+2} = \mathbf{K}^n = \mathbf{E}_1^{n+2} + \cdots + \mathbf{E}_{\alpha_{n+2}}^{n+2},$$

where $F(\mathbf{E}_\lambda^{n+2}) \subset \mathbf{K}^n$. Since the original systems $(\mathbf{r}, R)$ and $(\mathbf{r}', R')$ are L^*-equivalent, it follows from the addendum to theorem 2 that the augmented systems are L-equivalent, provided α_{n+2} is suitably chosen, which we assume to be the case. Therefore, by what we have already proved, the augmented system $(\mathbf{r}', R')$ is a natural system for a complex $\mathbf{K}_1^{n+2}$, which has the same nucleus, and hence the same $(n + 1)$-group as $\mathbf{K}^{n+2}$. It follows from the definition in §3 above that $\mathbf{K}^{n+2}$, and hence $\mathbf{K}_1^{n+2}$ has the same $(n + 1)$-group as $\mathbf{K}^n$. In general, if $(\mathbf{r}, R)$ is a natural system for a complex $\mathbf{K}^n$, then $(\mathbf{r}^p, \cdots, \mathbf{r}^3, R)$ $(3 \leqq p \leqq n)$ is a natural system for $\mathbf{K}^p$. Therefore the original system $(\mathbf{r}', R')$ is a natural system for $\mathbf{K}_1^n$. Since $\alpha'_{n+1} = 0$ the complex $\mathbf{K}_1^n$ is obtained from $\mathbf{K}_1^{n+2}$ by removing the interiors of the $(n + 2)$-cells, and hence has the same $(n + 1)$-group as $\mathbf{K}_1^{n+2}$. Therefore $\mathbf{K}_1^n$ has the same $(n + 1)$-group as $\mathbf{K}^n$, and the proof is complete.

If $X = L$ the order[40] of the deformation $D(\mathbf{K}^n) = \mathbf{K}_1^n$, in theorem 9, need not exceed $n + 1$. Also the formal deformations involved in theorem 5 are of order $n + 1$ at most. Therefore we have the corollary to theorems 5, 8 and 9:

COROLLARY. *If two complexes $\mathbf{K}_0^n$ and $\mathbf{K}^n$ have the same nucleus and either $n = 3$, or $n > 3$ and $\pi_r(\mathbf{K}^n) = 0$ for $r = 2, \cdots, n - 2$, then $\mathbf{K}^n = D(\mathbf{K}_0^n)$, where the order of D does not exceed $n + 1$.*

For let $(\mathbf{r}, R)$ and $(\mathbf{r}', R')$ be natural systems for $\mathbf{K}_0^n$ and $\mathbf{K}^n$ respectively. By theorem 8, $(\mathbf{r}', R')$ and $(\mathbf{r}, R)$ are L-equivalent. It follows from theorem 9 that $(\mathbf{r}', R')$ is a natural system for a complex $\mathbf{K}_1^n = D_1(\mathbf{K}_0^n)$, and from theorem 5, that $\mathbf{K}^n = D_2(\mathbf{K}_1^n)$, where D_1 and D_2 are of order $n + 1$ at most.

In $\pi_1(\mathbf{K}^n)$ satisfies the condition explained in S.S., §11, it follows from S.S., theorem 21, that the corollary remains valid with "have the same nucleus" replaced by "are of the same homotopy type." For example, an n-dimensional simplicial complex, which is an absolute retract, expands into a collapsible complex of at most $n + 1$ dimensions.

5. In this section we prove the theorem on lens spaces, which was announced in §1. We first prove a lemma concerning the ring of a cyclic group. Let $\mathfrak{G}$ be a cyclic group of order m, let $\mathfrak{R}$ be its group ring, and let $\sigma = 1 + \cdots + x^{m-1}$, where x generates $\mathfrak{G}$. We shall describe an element $\alpha \in \mathfrak{R}$ as *regular* mod. σ if, and only if, there is an element $\beta \in \mathfrak{R}$, such that

$$\alpha\beta = 1 + k\sigma, \tag{5.1}$$

for some value of k. Since $x^p\sigma = \sigma$ it follows that $\lambda\sigma = S(\lambda)\sigma$, for any $\lambda \in \mathfrak{R}$,

[40] By the order of a formal deformation is meant the maximum dimensionality of the cells introduced or removed.

where $S(\lambda)$ denotes the algebraic sum of the coefficients in λ. Therefore we assume that k, in (5.1), is an integer. The lemma is:

LEMMA 6. *If α is regular* mod. σ, *and if $S(\alpha) = \pm 1$, then α is regular.*

Writing $S(\alpha) = \epsilon$ we have $\alpha(\beta - \epsilon k\sigma) = 1 + k\sigma - k\sigma = 1$.

We also recall that, if $\lambda(x^q - 1) = 0$, where q is prime to m, then $\lambda \equiv 0$ (mod. σ), meaning that $\lambda = k\sigma$, for some value of k. If $\lambda(x^q - 1)^n = 0$, where $n > 1$, we have $\lambda(x^q - 1)^{n-1} = k\sigma$, whence $km\sigma = k\sigma^2 = \lambda(x^q - 1)^{n-1}\sigma = 0$. It follows from induction on n that, if $\lambda(x^q - 1)^n = 0$, for any $n > 0$, then $\lambda \equiv 0$ (mod. σ).

THEOREM 10. *Two lens spaces of types (m, q) and (m, r) are of the same homotopy type if, and only if,*[41] $r \equiv \pm l^2 q$ (mod. m), *for some value of l.*

Let M_1^3 be a lens space of type (m, q), which we assume to be represented by a lens model of this type. Then $(\mathbf{r}, R)$ is a natural system for M_1^3, where $\mathbf{r}$ consists of the matrix $\mathbf{r}^3 = \| x^q - 1 \|$, and R consists of a single generator x, and the relation $x^m = 1$. The corresponding incidence matrices are

$$\| x^q - 1 \|, \qquad \| \sigma \|, \qquad \| x - 1 \|.$$

If M_1^3 is of the same homotopy type as a lens space of type (m, r) it follows from theorem 8 and 6, and lemma 5, that

$$(x^{ql} - 1)(x^l - 1) = \alpha(x^r - 1)(x - 1), \tag{5.2}$$

for some l which is prime to m, and some regular element $\alpha \in \mathfrak{R}$. Let $\sigma_n = 1 + \cdots + x^{n-1}$ $(1 \leqq n \leqq m)$. Then $x^n - 1 = (x - 1)\sigma_n$, and it follows from (5.2) and a previous observation that

$$\sigma_{ql}\sigma_l \equiv \alpha\sigma_r \qquad \text{(mod. } \sigma).$$

On writing $x = 1$ we have $ql^2 \equiv \pm r$ (mod. m), since $S(\alpha) = \pm 1$, the element α being regular.

Conversely, let

$$r \equiv \pm l^2 q \qquad \text{(mod. } m). \tag{5.3}$$

If a and b are given integers and if $| a, m | = 1$, where $| c, d |$ denotes the positive H.C.F. of c and d, there is an element $\xi \in \mathfrak{R}$ such that[42] $\xi(x^a - 1) = x^b - 1$. If also $| b, m | = 1$ there is an η such that $\eta(x^b - 1) = x^a - 1$. Then

$$\begin{aligned} x^b - 1 &= \xi(x^a - 1) \\ &= \xi\eta(x^b - 1), \end{aligned}$$

whence $\xi\eta \equiv 1$ (mod. σ). Therefore $| b, m | = 1$ implies that ξ is regular mod. σ. Since $| r, m | = 1$ there is a $\xi \in \mathfrak{R}$, such that

$$\xi(x^r - 1) = x^{ql} - 1. \tag{5.4}$$

[41] Of course the necessity of this condition follows from the standard theorems on the Kronecker index and the fact that, if two manifolds are of the same homotopy type, then each is the image of the other in a map of degree 1.

[42] See p. 107 of Reidemeister, Abh. Hamb. Sem., 11 (1935), 102-9.

Also it follows from (5.3) that $| ql, m | = | l, m | = 1$. Therefore ξ is regular mod. σ, and so is $\sigma_l(x - 1) = x^l - 1$. Therefore $\beta = \xi\sigma_l$ is regular mod. σ. It follows from (5.4) that

$$\xi\sigma_r \equiv \sigma_{ql} \qquad (\text{mod. } \sigma),$$

whence $rS(\xi) \equiv ql$ (mod. m). Since $| ql, m | = 1$ it follows from (5.3) that $lS(\xi) \equiv \pm 1$ (mod. m), whence $S(\beta) = lS(\xi) = \pm 1 + tm$, for some integral value of t. Therefore $S(\alpha) = \pm 1$, where $\alpha = \beta - t\sigma$. Since β is regular mod. σ, so is α, and it follows from lemma 6 that α is regular. Since $(\beta - \alpha)(x^r - 1) = t\sigma(x^r - 1) = 0$, it follows from (5.4) that

$$\begin{aligned} \alpha(x^r - 1) &= \beta(x^r - 1) \\ &= \xi(x^r - 1)\sigma_l \qquad (5.5) \\ &= (x^{ql} - 1)\sigma_l . \end{aligned}$$

Since α is regular the natural system $(\| x^r - 1 \|, R)$, for a lens space M_2^3 of type (m, r), is L^*-equivalent to the system $\{\| \alpha(x^r - 1) \|, R\}$. By theorem 9 there is therefore a membrane complex $\mathbf{K}_2^3$, of the same homotopy type as M_2^3, for which $\{\| \alpha(x^r - 1) \|, R\}$ is a natural system. Therefore the theorem will follow from (5.5) and theorem 5 when we have shown that there is a membrane complex $\mathbf{K}_1^3$, of the same homotopy type as M_1^3, for which $\{\| (x^{ql} - 1)\sigma_l \|, R\}$ is a natural system. It will therefore follow from theorem 9 when we have shown that the system $(\| x^q - 1 \|, R)$ is L^*-equivalent to $\{\| (x^{ql} - 1)\sigma_l \|, R\}$ under some automorphism of $\mathfrak{G}_R$. Actually we shall show that the two systems are L-equivalent under the automorphism in which x corresponds to x^l.

Starting with $(\| x^q - 1 \|, R)$ we extend R to a system R' by the addition of a new generator y and the new relation $yx^{-l'} = 1$, where $ll' \equiv 1$ (mod. m). Such an l' exists since $| l, m | = 1$. At the same time we transform $\| x^q - 1 \|$ into the matrix $\| x^q - 1, 0 \|$. Let $x_1 = yx^l$, $y_1 = x$. Since $x = y_1$, $y = x_1 y_1^{-l}$, it follows that x_1, y_1 also generate the free group $\mathfrak{F}_2$, which is freely generated x, y, whence the transformation $x \to yx^l$, $y \to x$ determines an automorphism of $\mathfrak{F}_2$. By theorem 7, the system $(\| x^q - 1, 0 \|, R')$ is L-equivalent to $\{\| (yx^l)^q - 1, 0 \|, R''\}$ under the isomorphism given by $x \to yx^l$, $y \to x$, where R'' is generated by x, y, subject to the relations

$$(5.6) \qquad (yx^l)^m = 1, \qquad x(yx^l)^{-l'} = 1.$$

I say that the relations (5.6) can be reduced to $x^m = 1$ by transformations of the form (4.5), followed by one of the form T_2. For, since a relation of the form $x^a x_1^b$ may be replaced by $x_1^b x^a$, we may, for the purposes of the reduction, allow x^a to commute with x_1^b. Let us write

$$A = mx_1, \qquad B = x - l'x_1,$$

where $x_1 = yx^l$ and the addition is commutative. Then

$$l'A + mB = mx, \qquad kA + lB = lx - x_1,$$

where $km = ll' - 1$. Since $ll' - km = 1$ the transformation given by

$$A' = l'A + mB, \qquad B' = kA + lB$$

is the resultant of a sequence of elementary transformations of the form $(A, B) \to (A \pm B, B)$ or $(A, B \pm A)$. Therefore a similar sequence of the transformations (4.5), punctuated with suitable transformations of the form $x^a x_1^b \to x_1^b x^a$, will transform the relations (5.6) into the relations $x^m = 1$, $x^l x_1^{-1} = y^{-1} = 1$. On discarding the generator y and the relation $y^{-1} = 1$, it follows that the system $(\| x^q - 1 \|, R)$ is L-equivalent to $\{\| x^{ql} - 1)\lambda \|, R\}$, under the automorphism given by $x \to x^l$, where λ is the factor due to the transformation of the relations (5.6) into $x^m = 1$, $y^{-1} = 1$. It follows from theorem 6 that the sets of matrices

$$\| x^{ql} - 1 \|, \qquad \| \sigma \|, \qquad \| x^l - 1 \|$$

and

$$\| (x^{ql} - 1)\lambda \|, \qquad \| \sigma \|, \qquad \| x - 1 \|$$

are L-equivalent, and from lemma 5 that

$$\begin{aligned}(x^{ql} - 1)(x - 1)\lambda &= \pm x^p (x^{ql} - 1)(x^l - 1) \\ &= \pm x^p (x^{ql} - 1)(x - 1)\sigma_l ,\end{aligned}$$

for some value of p. Therefore

$$(x^{ql} - 1)\lambda = \pm x^p (x^{ql} - 1)\sigma_l .$$

On multiplying $(x^{ql} - 1)\lambda$ by $\pm x^{-p}$ it follows that $(\| x^q - 1 \|, R)$ is L-equivalent under the automorphism $x \to x^l$ to $\| (x^{ql} - 1)\sigma_l , R \|$, and the proof is complete.

6. We conclude with some notes. The sections to which they primarily refer are indicated by Roman numerals.

II$_1$. Let $\mathfrak{G}_0$ be the free product of an infinite cyclic group, generated by x, and a cyclic group of order two, generated by y. Let $\mathfrak{R}$ be the group ring of $\mathfrak{G}_0$, and let the group $\mathfrak{G}$, in terms of which elementary matrices are defined, consist of the elements $\pm x$, where $x \in \mathfrak{G}_0$. Finally let $\lambda = x(1 + y)$, $\mu = 1 - y$, where 1 stands for the unit element in $\mathfrak{R}$. Then $\lambda\mu = 0$, $\mu\lambda \neq 0$, and it may be verified that

$$\left\| \begin{matrix} 1 - \mu\lambda & 0 \\ 0 & 1 \end{matrix} \right\| = \mathbf{a}\mathbf{b}\mathbf{a}^{-1}\mathbf{b}^{-1},$$

where

$$\mathbf{a}^{\epsilon} = \left\| \begin{matrix} 1 & 0 \\ \epsilon\lambda & 1 \end{matrix} \right\|, \qquad \mathbf{b}^{\epsilon} = \left\| \begin{matrix} 1 & \epsilon\mu \\ 0 & 1 \end{matrix} \right\| \; (\epsilon = \pm 1).$$

Therefore an expansion of a matrix may be elementary, although the original matrix is not.

II_2. *Concerning the condition* (2.3). First notice that this is a consequence of the, apparently, less restrictive condition in which $\mathbf{a}$ and $\mathbf{b}$ are to be elementary matrices. For if the matrices $\begin{Vmatrix} 0 & \mathbf{e}_k \\ 0 & 0 \end{Vmatrix}$ and $\begin{Vmatrix} 0 & \mathbf{e}_l \\ 0 & 0 \end{Vmatrix}$ have p rows and q columns, we can border them with p final rows, and q initial columns of zeros, at the same time replacing given regular matrices, $\mathbf{a}$ and $\mathbf{b}$, by the elementary matrices $\begin{Vmatrix} \mathbf{a} & 0 \\ 0 & \mathbf{a}^{-1} \end{Vmatrix}$ and $\begin{Vmatrix} \mathbf{b}^{-1} & 0 \\ 0 & \mathbf{b} \end{Vmatrix}$.

This condition is also satisfied if $\mathfrak{R}$ is a ring in which $\mathbf{ab} = \mathbf{e}_r$ implies that $\mathbf{a}$ is regular, where $\mathbf{a}$ is a square matrix of degree r. The following example shows that it is not always satisfied. Let $\mathfrak{R}_0$ be the free ring, freely generated by a, b, p, q, whose elements are finite sums of words in a, b, p, q. Let $\mathfrak{R}$ be the ring which is obtained from $\mathfrak{R}_0$ by adding a unit element e, and the relations

$$ex = xe = x \qquad\qquad (x = a, v, p, q \text{ or } e)$$

$$ap + bq = pa = qb = e, \qquad pb = qa = 0.$$

The ring $\mathfrak{R}$ does not reduce to the zero element alone. For any formal polynomial in e, a, b, p, q can be reduced by means of these relations, together with the simplifications $mx + nx = (m + n)x$ $(m, n = 0, \pm 1, \pm 2, \cdots)$, $x + 0y = x$ and $xzy + xz'y = x(z + z')y$, to a unique normal form, in which the letters e, a, b, p, q occur a minimum number of times.[43] This can be proved by enumerating a set of "elementary reductions," which, with their inverses, generate the transformations by which any two expressions for a given element of $\mathfrak{R}$ can be interchanged. One can then show that the condition for "random reduction" is satisfied. That is to say, if two reductions r_1 and r_2 are applicable to a polynomial P, then $t_1r_1(P) = t_2r_2(P)$, where t_i either reduces $r_i(P)$ or leaves it unaltered. The expression e cannot be reduced, whence $e \neq 0$ in $\mathfrak{R}$.

Let

$$\mathbf{c} = \begin{Vmatrix} p & 0 & 0 \\ q & 0 & 0 \\ 0 & a & b \end{Vmatrix}, \qquad \mathbf{c}' = \begin{Vmatrix} a & b & 0 \\ 0 & 0 & p \\ 0 & 0 & q \end{Vmatrix}.$$

Then it follows by direct calculation that $\mathbf{cc}' = \mathbf{c}'\mathbf{c} = \mathbf{e}_3$ and that

$$\mathbf{c}\begin{Vmatrix} 0 & 0 & e \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix} = \begin{Vmatrix} 0 & 0 & p \\ 0 & 0 & q \\ 0 & 0 & 0 \end{Vmatrix} = \begin{Vmatrix} 0 & e & 0 \\ 0 & 0 & e \\ 0 & 0 & 0 \end{Vmatrix}\mathbf{c}'.$$

[43] E.g. a is to be counted once, not six times, in $6a$.

Notice that **c** is even an elementary $\mathfrak{G}$-matrix, where $\mathfrak{G}$ consists of the single element e. For it is transformed into the elementary matrix

$$\begin{Vmatrix} 0 & -e & 0 \\ 0 & 0 & -e \\ e & 0 & 0 \end{Vmatrix},$$

as follows

$$\rho_3 \to \rho_3 + a\rho_1 + b\rho_2\,; \qquad \kappa_2\,,\, \kappa_3 \to \kappa_2 - \kappa_1 a,\, \kappa_3 - \kappa_1 b;$$

$$\rho_1\,,\, \rho_2 \to \rho_1 - p\rho_3\,,\, \rho_2 - q\rho_3\,,$$

where ρ_i and κ_i denote the i^{th} row and the i^{th} column of a given matrix.

III_1. Two spaces X and Y are of the same homotopy type if, and only if, there are maps $f(X) \subset Y$ and $g(Y) \subset X$ such that $gf(X) \subset X$ and $fg(Y) \subset Y$ are each homotopic to the identity. We shall describe f as a *homotopy isomorphism.* Assuming that X and Y are arcwise connected, a homotopy isomorphism $f(X) \subset Y$ determines a unique class of isomorphisms $\psi\{\pi_1(X)\} = \pi_1(Y)$, any one of which is given by $\psi = \psi_0 h$, where ψ_0 is a particular isomorphism in the class, and h is an inner automorphism of $\pi_1(X)$.

III_2. The theory of nuclei and m-groups can also be developed as follows. Let $\mathbf{E}^n$ now stand for a non-singular cell, bounded by a given map $f(S^{n-1}) \subset X$, where X is any topological space. We recall the definition of $X + \mathbf{E}^n$, given in a paper entitled "On adding relations to homotopy groups" which has been submitted to the Annals of Mathematics, and which will be referred to as R.H. Let $S^{n-1} = \dot{E}^n$, where E^n is an n-element which does not meet X, and let $\mathbf{E}^n$ be the interior of E^n. Then $X + \mathbf{E}^n$ consists of X and $\mathbf{E}^n$, each with their own topology, fastened together by the condition that an infinite sequence of points $p_1\,, p_2\,, \cdots \subset \mathbf{E}^n \subset X + \mathbf{E}^n$ shall converge to $q \in f(\dot{E}^n)$ if, and only if, the limit points of $p_1\,, p_2\,, \cdots \subset \mathbf{E}^n$ are all contained in $f^{-1}(q) \subset \dot{E}^n$. In other words we identify each point $p \in \dot{E}^n$ with $f(p) \in X$, without identifying any two points in $\mathbf{E}^n$. A map $f(S^n) \subset X + \mathbf{E}^n$ may be described as *simple* in $\mathbf{E}^n$ if, and only if, it is a homeomorphism throughout $f^{-1}(U)$, where U is some relatively open set in $\mathbf{E}^n$. Then a membrane complex $\mathbf{K}^n$, and the nucleus and m-group of $\mathbf{K}^n$, may be defined as in §3 above. Many of the theorems in S.S., notably theorems 11, 14 and 17, remain valid when reinterpreted in terms of these definitions. Theorem 11, for example, may be proved as in §3 above. In theorem 17 one may take the analogue of a simplicial map to be a map $f(\mathbf{K}_0) \subset \mathbf{K}_1$, such that $f(\mathbf{K}_0^p) \subset \mathbf{K}_1^p$, for each $p = 0, 1, \cdots$.

Let $\mathbf{E}^n$ be a non-singular cell, which is bounded by a simplicial map $f(S^{n-1})$ in a simplicial complex K. Then it is to be expected that $\bar{\mathbf{E}}^n$, the closure of $\mathbf{E}^n$ in $K + \mathbf{E}^n$, can be triangulated.[44] Let us assume that $\bar{\mathbf{E}}^n$ can be triangulated to form a simplicial complex C^n, such that $C^n - \mathbf{A}^n$ contracts into the sub-

[44] It seems probable that $\bar{\mathbf{E}}^n$ is homeomorphic to a simple membrane, as defined in S.S.

complex covering $F(\mathbf{E}^n) = f(S^{n-1})$, where $\mathbf{A}^n$ is some open simplex of C^n, whose closure is in the (open) cell $\mathbf{E}^n$. Then it is easily shown that any membrane complex can be transformed into a polyhedron by a series of trivial deformations, as defined in §3 above; also that two simplicial complexes have the same nucleus, or m-group, in the sense of S.S., if they have the same nucleus, or m-group, when treated as membrane complexes. If the above assumption is correct it follows that the introduction of membrane complexes does not involve anything essentially new. But, in some problems, the freedom from the combinatorial limitations of S.S. is likely to be an advantage.

III$_3$. In this note we indicate how the incidence matrices for a membrane complex can be defined in terms of R.H. Let $\mathbf{E}_i^q$ $(q = 0, \cdots, n; i = 1, \cdots, m_q)$ be the cells of a membrane complex $\mathbf{K}^n$, which may be defined either as in §3 or as in III$_2$ above. Let $F(\mathbf{E}_i^p)$ be joined to $\mathbf{E}_1^0$ by a segment in $\mathbf{K}^{p-1}$ $(p > 2)$, so as to determine an element $\alpha_i^{p-1} \in \pi_{p-1}(\mathbf{K}^{p-1})$. Then $\mathbf{r}^p$ is the matrix of the coefficients in

$$\psi(\alpha_\lambda^{p-1}) = \sum_{i=1}^{m_{p-1}} r_{\lambda i}^p e_i^{p-1}, \tag{6.1}$$

where e_i^{p-1} is the basis element of the modulus M, in R.H., which corresponds to $\mathbf{E}_i^{p-1}$. Theorem 5 above, and lemma 3 in R.H., enable us to describe the indeterminacy in the description of $\mathbf{K}^n$ by means of a natural system $(\mathbf{r}, R)$. The nucleus of $\mathbf{K}^3$ is uniquely determined by R and $\mathbf{r}^3$. When $\mathbf{K}^{p-1}$ is given, for $p > 3$, the elements $\alpha_1^{p-1}, \cdots, \alpha_{m_p}^{p-1}$ determine $\mathbf{K}^p$ up to a trivial deformation rel. $\mathbf{K}^{p-1}$, and $\mathbf{r}^p$ is given by (6.1). In passing from the description of $\mathbf{E}_1^p, \cdots, \mathbf{E}_{m_p}^p$ by means of $\alpha_1^{p-1}, \cdots, \alpha_{m_p}^{p-1}$, to their description by means of $\mathbf{r}^p$, we throw away the sub-group $\overset{0}{\pi}_{p-1} \subset \pi_{p-1}(\mathbf{K}^{p-1})$, whose elements have representative maps $f(S^{p-1}) \subset \mathbf{K}^{p-2}$.

For example, let $\mathbf{K}_i^4 = \mathbf{E}_i^2 + \mathbf{E}_i^4$ $(i = 0, 1)$, where $F(\mathbf{E}_i^2) = \mathbf{E}_i^0 = \mathbf{K}_i^0$, and $\mathbf{E}_i^4$ is bounded by a map $F(\mathbf{E}_i^4) \subset \mathbf{E}_i^2$, whose Hopf invariant[45] is γ_i. Then $\mathbf{K}_0^4$ and $\mathbf{K}_1^4$ have the same natural system, namely the empty system in which R has no generators and a single vacuous relation, and $m_3 = 0$, $m_4 = 1$. It follows from what is known concerning[46] $\pi_3(S^2)$, and S.S., theorem 18, that $\pi_3(\mathbf{K}_i^4)$ is cyclic of order $|\gamma_i|$. Therefore $\mathbf{K}_0^4$ and $\mathbf{K}_1^4$ are not of the same homotopy type if $\gamma_1 \neq \pm\gamma_0$. If $\gamma_0 = 1, \gamma_1 = 0$ we may take $\mathbf{K}_0^4$ to be the complex projective plane and $\mathbf{K}_1^4$ to consist of a 4-sphere, $\mathbf{E}_1^4$, and a 2-sphere, $\mathbf{E}_1^2$, with a single common point $\mathbf{E}_1^0$. In this case their homotopy types are also distinguished by the fact that the product[47] of a basic 2-dimensional co-cycle, or surface integral, with itself is co-homologous to zero in the case of $\mathbf{K}_1^4$, and not in the case of $\mathbf{K}_0^4$. Can $\mathbf{K}_0^4$ and $\mathbf{K}_1^4$ be distinguished from each other in this way for arbitrary values of γ_0 and $\gamma_1 \neq \pm\gamma_0$?

[45] See H. Hopf, Math. Annalen, 104 (1931), 637–65.

[46] See H. Freudenthal, Compositio Math., 5 (1938), 299–314.

[47] See Hassler Whitney, Proc. Nat. Academy of Sciences, 23 (1937), 285–91, and papers by E. Čech and J. W. Alexander, which are referred to there.

III$_4$. *Note on the calculation of* $\pi_r(K)$ $(r > 1)$. We shall say that a group $\mathfrak{G}$, with a countable set of elements, has been *calculated* when it is given in terms of generators and relations, which may be infinite in number, but which must be enumerated. That is to say, a process must be described by which all the generators and relations, may be successively constructed.[48] We shall say that $\mathfrak{G}$ has been *calculated effectively* when it has been calculated, and when a finite algorithm has been provided by which one can tell whether or no two products of the generators represent the same element of $\mathfrak{G}$.

We now show how $\pi_r(K)$ may be calculated for any $r \geqq 1$, where K is a simplicial complex, which may be infinite but locally finite. The group $\pi_1(K)$ can be calculated in various ways, as in §3 for example, and we therefore take $r > 1$. Let $a_1, a_2, \cdots$ be a given, enumerable sequence of vertices and let us define the *weight* of a given symbolic r-simplex $A^r = a_{i_0} \cdots a_{i_r}$ $(i = 1, 2, \cdots)$ as $w(A^r) = i_0 + \cdots + i_r$. We order all the r-simplexes by placing A_1^r before A_2^r if $w(A_1^r) < w(A_2^r)$ and ordering the r-simplexes with a given weight lexicographically. Similarly, we order the purely r-dimensional complexes, $P^r = A_{i_0}^r + \cdots + A_{i_q}^r$, in terms of the r-simplexes. By a *special* r-element we shall mean one which is derived from an r-simplex by reiterating the two operations of stellar sub-division, and addition $E_1^r + E_2^r$, where E_1^r and E_2^r are special r-elements which meet in an $(r - 1)$-simplex on the boundary of both. One can find out, by a finite process of trial and error, whether or no a given complex P^r is a special r-element. Therefore the special r-elements are enumerated in the form of a sub-sequence of the purely r-dimensional complexes. Assuming, as we shall do, that the vertices of K are in a given order, one can enumerate the simplicial maps $f(E^r) \subset K$, where E^r is a given r-element, and hence the aggregate of simplicial maps $f_{\lambda i} = f_{\lambda i}(E_i^r) \subset K$ $(\lambda = 1, \cdots, n_i; i = 1, 2, \cdots)$, such that $f_{\lambda i}(\dot{E}_i^r)$ is a given vertex $p_0 \in K$, where $E_1^r, E_2^r, \cdots$ are the special r-elements. Any element of $\pi_r(K)$ is represented by such a map, and indeed by one in which E_i^r is a stellar sub-division of an r-simplex. Therefore the set of maps $f_{\lambda i}$ may be taken as the generators of a symbolic group,[49] which, with suitable relations, represents $\pi_r(K)$. For relations we take, first all relations of the form $f_{\lambda i} f_{\mu j} = f_{\rho k} = \cdots = f_{\sigma l}$, which are defined by adding the special r-elements E_i^r and E_j^r, in all possible ways, to form the special r-elements $E_k^r, \cdots, E_l^r$. By means of these relations, which we shall call the relations R, any product of the generators can be transformed into a single generator. Therefore, in addition to the relations R, we only need a set by means of which any representative map, $f_{\rho i}$, of the element $1 \in \pi_r(K)$, can be transformed formally into the constant map f_{1i}, say, together with the relations $f_{1i} = 1$ $(i = 1, 2, \cdots)$. Let A be any internal simplex in E_i^r, let $B = f_\lambda(A)$, for a given value of λ, and let bB be a simplex in K such that, if $a_j A \subset E_i^r$, then $f_\lambda(a_j A) \subset bB$ (the vertex b may belong to B, in which case $bB = B$). Let σ be the elementary

[48] E.g. a process for calculating the n^{th} prime number.

[49] Cf. III$_5$ below. Here we write $\pi_r(K)$ with multiplication, even though $r > 1$.

subdivision (A, a_k), where a_k is the first of the vertices $a_1, a_2, \cdots$ which is not in E_i^r, let $E_j^r = \sigma E_i^r$, and let $f_\mu(E_j^r) \subset K$ be the (simplicial) map which is given by $f_\mu(a_l) = f_\lambda(a_l), f_\mu(a_k) = b$, where a_l is any vertex in E_i^r. Then E_j^r is a special r-element, and $f_{\mu j}$ is one of the maps defined above. Let R′ be the set of all relations $f_{1i} = 1, f_{\lambda i} = f_{\mu j}$, where $f_{\lambda i}$ and $f_{\mu j}$ are related in this way. Then, using S.S., Lemma 2 and Theorem 6, it is not difficult to prove that the symbolic group defined by the generators $f_{\lambda i}$ $(\lambda, i = 1, 2, \cdots)$ and the relations R, R′ represents $\pi_r(K)$. The relations R, R′ can obviously be enumerated, whence $\pi_r(K)$ can be calculated for any value of $r = 1, 2, \cdots$. In general it cannot be calculated effectively, even when $r = 1$ and K is finite. Thus if K is infinite $\pi_r(K)$ is on the same logical footing whether $r = 1$ or $r > 1$. If K is finite there is the distinction that $\pi_1(K)$, unlike $\pi_r(K)$ $(r > 1)$, can always be represented by a finite system of generators and relations.

III_5. By a symbolic group we mean a group which is originally defined in terms of given generators and relations. An element in such a group is a class[60] of products of the generators and their inverses, and is to be distinguished from an individual product. However we shall use the same symbol to denote a product of generators and the element which it represents. If W_1 and W_2 are two products, then $W_1 = W_2$ is to be interpreted as an equation between elements of a group, unless the contrary is obviously implied by the context, as it is in (4.1), for example.

V. This note contains a simpler example than the one in §5 above, showing that two complexes of the same homotopy type need not have the same nucleus. Let $\mathbf{K}^p$ $(p = 0, 1, 2)$ consist of a p-cell $\mathbf{E}^p$, such that $F(\mathbf{E}^2) = 5\mathbf{E}^1$. Thus the single generator x, and the relation $x^5 = 1$, form a natural system R, for $\pi_1(\mathbf{K}^2)$, and $\tilde{\mathbf{K}}^2$ consists of five 2-cells $x^m\tilde{\mathbf{E}}^2$ $(m = 0, \cdots, 4)$, each bounded by $\sigma\tilde{\mathbf{E}}^1$ $(\sigma = 1 + x + \cdots + x^4)$, where $\tilde{\mathbf{E}}^1 \subset \tilde{\mathbf{K}}^1$ covers $\mathbf{E}^1 \subset \mathbf{K}^1$. A chain $\lambda\tilde{\mathbf{E}}^2$ $(\lambda \in \mathfrak{R})$ is a cycle if, and only if $\lambda\sigma = 0$, and by Hurewicz's theorem any 2-cycle in $\tilde{\mathbf{K}}^2$ is realized by some simplicial map $f(S^2) \subset \tilde{\mathbf{K}}^2$. Therefore $\{\| (x - 1)^2 \|, R\}$ is a natural system for a complex $\mathbf{K}^3 = \mathbf{K}^2 + \mathbf{E}^3$, such that $\partial\tilde{\mathbf{E}}^3 = (x - 1)^2\tilde{\mathbf{E}}^2$. It may be verified that $x^5 = 1$ implies $(x^3 + x^2 - 1)(x^4 + x - 1) = 1$, whence $\alpha = x^3 + x^2 - 1$ is a regular element of $\mathfrak{R}$. Therefore the system $\{\| \alpha(x - 1)^2 \|, R\}$ is L^*-equivalent to $\{\| (x - 1)^2 \|, R\}$, and by theorem 9 there is a complex $\mathbf{K}_1^3$, of the same homotopy type as $\mathbf{K}^3$, for which $\{\| \alpha(x - 1)^2 \|, R\}$ is a natural system. Let us assume that $\mathbf{K}^3$ and $\mathbf{K}_1^3$ have the same nucleus. Then it follows from theorems 8 and 6 that the sets of incidence matrices

$$\| (x^l - 1)^2 \|, \qquad \| \sigma \|, \qquad \| x^l - 1 \|$$

$$\| \alpha(x - 1)^2 \|, \qquad \| \sigma \|, \qquad \| x - 1 \|$$

[60] Cf. K. Reidemeister, Einführung in die kombinatorische Topologie, Brunswick (1932), chap. 2.

are L-equivalent for some $l = 1, \cdots, 4$, and from lemma 5 that

$$\alpha(x-1)^3 = \pm x^p(x^l - 1)^3,$$

for some value of p. Since

$$x^p(x^l-1)^3 = -x^{p+3l}(x^{-l}-1)^3,$$

we may assume that $l = 1$ or 2. As explained in §5 above, $\lambda(x-1)^3 = 0$ implies $\lambda \equiv 0$ (mod. σ). Clearly $\alpha \not\equiv \pm x^p$ (mod. σ), whence $l = 2$. Therefore $\alpha \equiv \pm x^p(x+1)^3$ (mod. σ), whence, writing $x = 1$, we have $1 \equiv \pm 8$ (mod. 5), which is absurd. Therefore $\mathbf{K}^3$ and $\mathbf{K}_1^3$ do not have the same nucleus.

BALLIOL COLLEGE,
OXFORD.

INDEX

Volume 42 of the ANNALS OF MATHEMATICS consists of five numbers. The publication of the supplementary number is made possible by a grant from the American Philosophical Society.